PORTFOLIO RISK MEASUREMENT: THE ESTIMATION OF THE COVARIANCE OF STOCK RETURNS

By

Lan Liu

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF WARWICK, WARWICK BUSINESS SCHOOL
COVENTRY, UNITED KINGDOM
JUNE 2007

To my parents and my husband.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to take this opportunity to thank my advisor, Professor Stewart Hodges, to whom I am greatly indebted. Professor Hodges has been a great source of intellectual inspiration and also a great mentor who patiently guides me over the years.

I am thankful to my parents and my husband. Without their support and encouragement, it would be impossible to complete this thesis.

Finally I thank the Warwick Business School for providing Bursary support.

Lan Liu

June, 2007

Declaration

I declare that this thesis is submitted to the University of Warwick for the degree of Doctor of Philosophy in 2007. Except where acknowledged, the material contained in this thesis is my own work and has neither been previously published nor submitted elsewhere for the purpose of obtaining an academic degree.

Lan Liu
June, 2007

Abstract

A covariance matrix of asset returns plays an important role in modern portfolio analysis and risk management. Despite the recent interests in improving the estimation of a return covariance matrix, there remain many areas for further investigation. This thesis studies several issues related to obtaining a better estimation of the covariance matrix for the returns of a reasonably large number of stocks for portfolio risk management.

The thesis consists of five essays. The first essay, Chapter 3, provides a comprehensive analysis of both old and new covariance estimation methods and the standard comparison criteria. We use empirical data to compare their performances. We also examine the standard comparisons and find they provide limited information regarding the abilities of the covariance estimators in predicting portfolio variances. It therefore suggests that we need more powerful comparison criteria to assess covariance estimators.

The second and third essays, Chapter 4 and 5, are concerned with the alternative appraisal methods of return covariance estimators for portfolio risk management purposes. Chapter 4 introduces a portfolio distance measure based on eigen decomposition (*eigen-distance*) to compare two covariance estimators in terms of the most different portfolio variances they predict. The *eigen-distance* measures the ratio of the two extreme variance predictions under one covariance estimator for the portfolios that are constructed to have the same variances under the other covariance estimator. We show that the *eigen-distance* can be used to assess a risk measurement system as a whole, where any kind of the portfolios may need to be considered. Our simulation results show that it is a powerful measure to distinguish two covariance estimators even in small samples.

Chapter 5 proposes a θ measure to distinguish two similar estimated covariance matrices from the observed covariance matrix. θ is constructed based on the essential

difference of the two similar covariance matrices: the two extreme portfolios that are predicted to have the most different variances under these two matrices. We show that θ is very useful in evaluating refinements to covariance estimators, particularly a modest refinement, where the refined covariance matrix is close to the original matrix.

The last two essays, Chapter 6 and 7, are concerned with improving the best covariance estimators within the literature. Chapter 6 explores alternative Bayesian shrinkage methods that directly shrink the eigenvalues (and in one case the principal eigenvector) of the sample covariance matrix. We use simulations to compare the performance of these shrinkage estimators with the two best existing estimators, namely, the Ledoit and Wolf (2003a) estimator and the Jagannathan and Ma (2003) estimator using both RMSE and eigen-distance criteria. We find that our shrinkage estimators consistently out-perform the Ledoit and Wolf estimator. They also out-perform the Jagannathan and Ma estimator except in one case where they are not much worse off either.

Finally, Chapter 7 extends the analysis of Chapter 6, which is under an unchanging multivariate normal world, to consider implications of both fat-tails and time variation. We use a multivariate normal inverse Gaussian (MNIG) distribution to model the log returns of stock prices. This family of distributions has proven to fit the heavy tails observed in financial time series extremely well. For the time varying situation, we use a tractable mean reverting Ornstein-Uhlenbeck (OU) process to develop a new model to measure an interesting and economically motivated time varying structure where the risks remain unchanged but stocks migrate among different risk categories during their life circles. We find that our shrinkage methods are also useful in both situations and become even more important in the time varying case.

Chapter 1

Introduction

Many modern financial applications such as portfolio construction (Markowitz (1952)) and risk management, require estimates of the covariance matrix of asset returns. The finance literature in the past paid less attention to the estimation of covariance than expected returns and variance. This lack of attention is due to two factors. First, there was limited computing technology to practically handle large amount of cross-sectional information for covariance estimation.[1] Second, it was generally believed that in a mean-variance optimization process, compared to expected returns, covariance is more stable and causes fewer problems; hence it is less important to have good estimations for it. Recently, with development in both optimization and computational technologies and renewed interests in portfolio risk management, there has been increasing attention on covariance estimations.

[1]For a portfolio with n assets, there are $n(n-1)/2$ pairs of covariances. The number of covariances increases very rapidly as the number of securities in the portfolio rises towards any realistic level, which poses a serious problem for a limited computing technology. As an example, for a portfolio with 100 stocks, we need to estimate 100 returns and 100 variances but 4,950 covariances.

This thesis studies several issues related to obtaining a better estimation of the covariance matrix for the returns of a reasonably large number of stocks for portfolio risk management. To set the stage for our analysis, we first conduct a comprehensive examination of the existing covariance estimation methods and standard comparison criteria. Next, as part of exploring the best way to estimate a return covariance matrix, we propose robust alternative appraisal criteria that can be applied to measure a risk system as a whole, where the risk of any portfolio may need to be considered. Finally, we explore the improvements on the best covariance estimation methods within the literature.

1.1 The scope of the thesis

The thesis consists of five essays. The first essay is an empirical study on comparing the covariance estimation methods. We contribute to the literature by providing an up-to-date analysis of both old and new estimation methods (altogether eleven methods[2]). We compare these methods using the conventional comparison criteria, i.e., in terms of both their abilities to produce accurate pair-wise estimated covariance (the Root Mean Square Error (RMSE) measure) and the volatility of the minimum

[2]These eleven estimation methods are the sample historical method, the single-index covariance estimator, the single-index model with Blume-adjusted betas estimator, the single-index model with Vasicek-adjusted betas estimator, the multi-index industry factor estimator, the multi-index principal components factor estimator, and the over- all mean estimator, the RiskMetrics exponentially weighted moving average method, the Ledoit and Wolf (2003a) Bayesian shrinkage method, the Jagannathan and Ma (2003) estimator, and the random matrix filtering method.

variance portfolios (the MVP test).[3]

While we find some consistent results from both the RMSE and MVP test that confirm with the existing literature such as the sample historical covariance matrix estimator under-performs most covariance estimators, we find these two comparisons in general give different results as they measure different things. The RMSE is a statistical measure and does not provide any implications regarding the use of the estimated covariances. While the MVP test has an economic implication, it provides only limited information on the ability of the covariance estimators in predicting portfolio variances as it is based on only one special portfolio. This shows that we need to find a more powerful measure to compare the performance of alternative covariance estimators. For portfolio risk management purposes, we want a more robust assessment criterion where the risks of any portfolios can be measured. This leads to the second part of our thesis, where we propose robust new measures to compare the alternative estimates of the return covariance matrix.

Given a variety of alternative covariance estimation methods, a key question is how to best choose among them. We know that a superior return covariance estimator is based on both the quality of the estimator and the measurement criteria (Makridakis et al. (1982, 1993, 2000)). Using comparison criteria that do not fit the purpose of the use of a covariance matrix may produce very misleading results. However,

[3]See for example, Elton and Gruber (1973), Elton, Gruber and Urich (1978), Chan, Karceski and Lakonishok (1999), Ledoit and Wolf (2003a), Jagannathan and Ma (2003), and Elton, Gruber and Spitzer (2006).

the existing studies have not paid enough attention to the comparison criteria. We believe that as part of exploring the best way to estimate a return covariance matrix, it is necessary to consider a better appraisal method.

The second essay is therefore motivated by the need to search for a robust measure which not only produces consistent statistical and economic comparison results but also is suitable for portfolio risk management. We are particularly interested in the maximum errors of a risk system. More specifically, we want to know how far the portfolio variances estimated under one covariance estimator could differ from the variances predicted under another covariance estimator.

In Chapter 4, we introduce a portfolio-distance measure based on eigen decomposition (*eigen-distance*) to compare two covariance matrices in terms of the biggest differences between the portfolio variances they predict. More specifically, we measure that if we construct portfolios to have the same variances under one covariance matrix, how differently these portfolios will be under the second covariance matrix. Our *eigen-distance* equals the log ratio of the two most extreme portfolio variance predictions under the second covariance matrix. It therefore measures the biggest difference of the two covariance matrices in terms of predicting the variances of these portfolios. Geometrically, this *eigen-distance* can also be viewed as simply the difference of two matrices in terms of their most different length of all directions under the same coordinate system.

We show that our *eigen-distance* is suitable for applying to a risk measurement system as a whole, where any kind of the portfolios may need to be considered. It can be applied equally well to both the absolute variances of portfolios and the variances of their tracking errors against a benchmark. We prove that this new measure is a proper measure of distance. In addition, we use simulations to show that it is very powerful to distinguish two covariance estimators even in small samples. In summary, our *eigen-distance* is an operational and powerful new measure that can be applied to evaluate empirical covariance matrices of large portfolios.

The third essay is concerned with an appropriate measure to evaluate refinements to the covariance estimators. More specifically, we want to find a suitable method to compare a refined covariance matrix and the original covariance matrix when we have the observed covariance matrix. We are particularly interested in a modest refinement, where the refined covariance estimator is relatively close to the original one. We want to find out if the refinement makes the estimated covariance matrix closer to the observed covariance matrix than the original covariance matrix.

In Chapter 5, we introduce a new θ measure to assess which one of two similar covariance matrices is closer to an observed covariance matrix. θ is designed to measure how much the two covariance matrices differ and whether one matrix is a clear improvement of the other. It is based on the performance of the two extreme portfolios that are predicted to have the most different variances under the two estimated

covariance matrices, which is the essential difference of the two matrices. We find that the θ measure is more powerful than both the RMSE and eigen-distance measures in differentiating two similar covariance matrices.

Having considered the more robust alternative criteria to evaluate the covariance estimation methods, the third part of the thesis is concerned with improving the best covariance estimators within the literature. A number of recent studies have found that the Ledoit and Wolf (2003a) estimator, which shrinks the sample covariance matrix towards a structured single-index covariance estimator, and the Jagannathan and Ma (2003) estimator, which takes an equally-weighted average of the sample covariance matrix and the single-index covariance estimator, work better than many other covariance estimators.[4] We are interested in alternative covariance estimation methods that can out-perform these covariance estimators for portfolio risk management purposes.

In Chapter 6, we explore some Bayesian shrinkage estimators based on directly shrinking the eigenvalues (and in one case the principal eigenvector as well) of the sample covariance matrix. Daniel and Kass (2001) review the empirical Bayesian shrinkage estimators proposed in recent years. Several authors have focused on shrinking the eigenvalues to improve the sample covariance matrix. This is because the eigenvalues of the sample covariance matrix tend to be more dispersed than the eigenvalues of the population covariance matrix. Therefore an intuitively appealing approach is

[4]See for example, Ledoit and Wolf (2003a, 2003b, 2004), Jagannathan and Ma (2003).

to shrink the sample eigenvalues towards some central value. We impose minimum structures for the eigenvalues and adjust the sample eigenvalues of two successive sample covariance matrices in a way similar to the Blume (1971) adjustment on the estimated betas in the single-index model. This is inspired by the similar behavior of the sample eigenvalues and the estimated betas in that high eigenvalues (estimated betas) tend to have positive estimation errors and low eigenvalues (estimated betas) tend to have negative estimation errors.

We use simulations to compare the performance of our shrinkage covariance estimators with the two best existing covariance estimators using both the RMSE and eigen-distance criteria, and find that we have succeed to a considerable extent. Our estimators not only consistently beat the Ledoit and Wolf estimator, they also outperform the Jagannathan and Ma estimator by a considerable amount in most circumstances except in one case, where they are not much worse than the Jagannathan and Ma estimator either.

So far our simulations have been based on the assumption of an unchanging multivariate normal world. The final essay of the thesis extends the analysis of Chapter 6 to explore the implications of both fat tails and time variation. More specifically, in the first extension, we use a multivariate normal inverse Gaussian (MNIG) distribution to model the log returns of stock prices. This family of distributions has proven to fit the heavy tails observed in financial time series extremely well.

The second extension considers a more interesting and economically motivated time varying covariance structure where the general market risk characteristics are constant while stocks migrate among different risk categories during their life cycles. It is not obvious what kind of a multivariate GARCH model would be necessary to have this set of properties, therefore we develop a different kind of model. This original model employs a tractable mean reverting Ornstein-Uhlenbeck (OU) process to model a multivariate factor model where the factor loadings change over time while the factors remain constant. This gives us a time varying covariance structure where the cross-section of risk characteristics tends not to change but the identities of stocks do.

We study how the relative performance of various covariance estimators is affected by these more realistic covariance structures and what types of methods become more important under these circumstances. We find that our eigenvalue shrinkage estimators (and the shrinkage on the principal eigenvector) are still very useful in both situations, and become even more important in the time varying case. The time varying covariance structure raises many issues that need to be explored further. Our work in this essay has explored and contributed to the understanding of how things change in the time varying situation.

1.2 The organization of the thesis

Following the introduction, the rest of the thesis is organized in the following way.

Chapter 2 reviews the literature on the covariance estimation methods and the standard comparison criteria. Chapter 3 conducts a comprehensive empirical analysis of both old and new methods of estimating the return covariance matrix using the standard comparison criteria. We examine the limitations of these comparisons and find that for portfolio risk management purposes, we need more robust comparison criteria which we can use to assess the risk of any portfolios to compare two covariance estimators. In Chapter 4, we introduce a robust portfolio distance measure based on eigen decomposition (*eigen-distance*) to compare two covariance matrices in terms of the most different portfolio variances they can predict. We show that *eigen-distance* is a powerful measure that can be used to assess the risk of any portfolios and can be easily applied to compare the alternative estimates of the covariance matrix for the returns of a reasonably large number of stocks. Chapter 5 introduces a θ measure to evaluate a modest refinement on a covariance estimator. We show that θ is very useful to differentiate two similar covariance matrices from an observed covariance matrix. In Chapter 6, we explore alternative Bayesian shrinkage methods based on directly shrinking the eigenvalues (and principal eigenvector) to improve on the best covariance estimators within the literature. We use simulations to investigate their relative performance under a constant covariance structure and find our shrinkage

methods consistently out-perform the alternative estimators. Chapter 7 extends the analysis of Chapter 6 to consider the implications of both fat-tails and time variation on the relative performance of the various covariance estimators. We find that our shrinkage methods are still helpful under both circumstances if not more important. We also explore how things change in the time varying situation and propose an original tractable time varying covariance structure where stocks migrate among risk categories during their life cycles. Finally, Chapter 8 concludes and describes potential future research.

Chapter 2

Literature review

Estimation of the covariance matrix of asset returns plays an important role in both the theory and practice of modern portfolio analysis and financial risk management. Markowitz's (1952) mean-variance portfolio optimization theory shows that we can construct optimal portfolios if accurate estimation of expected returns, variance and covariance of every asset could be obtained. Following the work of Markowitz, numerous studies have been searching for methods that can provide the best estimates of the inputs required for this mean-variance analysis. More recently, DeMiguel, Garlappi and Uppal (2005) find that for many asset allocation problems, the large error in estimating moments of asset returns may overwhelm the gains from optimization. Therefore, while there has been considerable progress in the design of optimal portfolios, more energy needs to be devoted to improving the estimation of parameters for the moments of asset returns.

This chapter provides a brief review of the literature on estimating a return covariance matrix[1] and on comparing alternative covariance estimators.

2.1 Literature on estimating a covariance matrix

The literature on estimating a return covariance matrix is quite extensive. We focus our review on methods that use mostly historical stock return data but limited stock fundamental information such as the industry classifications in the estimation process. This means that we do not consider the more elaborate factor models such as those by Rosenberg (1974) or Roll and Ross (1980), which require a substantial amount of non-price data.

The traditional methods of estimating an empirical return covariance matrix can be broadly classified into three general types: the sample historical model, index (factor) models and average models.[2] Lately with the development in optimization and computational technologies, methods with more complex techniques are proposed.

2.1.1 Sample historical covariance matrix

Although the sample covariance matrix based on historical return data contains some useful information about future covariances, many studies find that it contains a lot

[1] A lot of studies estimate the correlation coefficient instead of the covariance of the stock returns. A correlation coefficient is closely related to covariance: $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$; and it also has a nice property that $|\rho_{ij}| \leq 1$ which provides convenience in analysis. There is some difference in estimating correlation and covariance. This thesis focuses on the estimation of covariances.

[2] See Elton, Gruber, Brown and Goetzmann (2007).

of noise and under-performs other methods as a forecast for the future covariance matrix.[3] The main problem with the sample covariance matrix is that it requires too many parameters for estimation. More specifically, if we have N stocks in a portfolio, we need to estimate $\frac{N(N-1)}{2}$ pairs of covariances for the covariance matrix. When the number of stocks (N) is of the same order as, or larger than, the number of historical returns per stock (T), the covariance matrix contains a lot of estimation errors.[4]

This problem is tackled from a number of aspects. One method is to impose some structure on the covariance matrix to reduce the estimation parameters. Although this introduces some specification errors in the estimation process, it could improve the overall performance of the estimated covariance matrix. Another method is to use high frequency data to increase the observations for estimation. Studies have found that using high frequency data often incurs other problems related to market microstructure. In addition, as portfolios are usually managed on a relatively long-term basis (as compared to stocks which may be traded on a more frequent minute or even second basis), we cannot rely on using high frequency data to improve the estimation of a covariance matrix for portfolio management purposes.

---

[3] See for example, Elton and Gruber (1973), Elton, Gruber and Urich (1978) and Eun and Resnick (1984).

[4] Pafka and Kondor (2004) refer this as the problem of the *curse of the dimensions.*

2.1.2 Index models

The most well known structural model is the Sharpe's single-index model (Sharpe (1963)), which assumes that stocks move together only because of their common responses to the market index. Empirical studies have found that the single-index covariance estimator outperforms the sample historical correlation matrix.[5] This indicates that a large part of the observed covariance structure between securities, not captured by the single-index model, represents random noise with respect to forecasting.

The use of the single-index model requires estimates of the beta for each stock included in a portfolio. Beta measures the sensitivity of a stock's return to the return of the market. Studies find that betas in the estimation period tend to be closer to one (which is the market beta) than the estimates obtained using the ordinary least square (OLS) from the historical data (Blume (1971,1975), Levy (1971)). As the single-index model is only as good as the estimates of betas, a number of techniques have been proposed to adjust betas' regression tendency. The two best known techniques are the Blume's (1971) linear regression adjustment and the Vasicek's (1973) Bayesian adjustment.

Blume (1971) regresses the estimated betas obtained from one historical period on the estimated betas obtained from a prior period and uses this regression to adjust

[5]See for example, Cohen and Pogue (1967), Elton and Gruber (1973), and Elton, Gruber, Brown and Goetzmann (2007).

betas for the estimation period. Vasicek (1973) on the other hand suggests that rather than adjusting all stocks with the same amount towards the average, the estimated betas should be adjusted according to the size of their uncertainty. Hence if there is a greater sampling error of a beta, there should be a greater adjustment on it to reflect the fact that the particular beta has a greater chance of having large differences from the average. Empirical evidence shows that both adjustments improve the estimation of a covariance matrix (Elton, Gruber and Urich (1978), Klemkosky and Martin (1975)). The evidence on the choice between the Blume and Vasicek adjustment is mixed, but the Vasicek adjustment seems to work slightly better.

As the sample historical covariance matrix can be regarded as obtained from a full factor model, while the single-index model estimator is from a one-factor model, the intuition is that a better estimator can be found between these two extreme models. Empirical evidence suggests that there are common influences beyond the market effect. The multi-index models attempt to capture the non-market factors that cause stocks to move together. The additional factors that have been suggested include the industry factors (King (1966)), the macroeconomic factors (Chen, Ross and Roll (1986)), the stock fundamental factors (Fama and French (1993)), and the statistical factors (principal components factors). Connor (1995) compares the explanatory power of using different types of factors for the U.S. equity returns. He finds that the statistical and fundamental factor models substantially outperform the

macroeconomic factor model, with fundamental factor model slightly outperforming the statistical factor model.

The disadvantage of multi-index models is that there is no consensus on either the nature or the number of the non-market factors (Connor and Korajczyk (1993)). This does not mean that none of the multi-index models may work well, but means that we do not know which model works well for a particular circumstance in advance. Studies have found that although adding more indices leads to a better explanation of the historical correlation matrix, it leads both to a poorer prediction of the future correlation matrix and to a lower return at specified risk levels for the out-of-sample efficient portfolios. This shows that adding *extra-market* indices picks up random noises rather than information in forecasting covariances.

2.1.3 Shrinkage models

Ledoit and Wolf (2003a) introduce a shrinkage method as an alternative way to impose a factor structure, which avoids specifying the number and nature of factors beyond the market factor. They take a weighted average of the sample covariance matrix and a single-index model estimator (as there is strong consensus regarding the market factor), and use the weight assigned to the single-index model estimator to control how much structure to impose. The authors find that the combined covariance estimator performs better than either estimator itself and other traditional estimators. In their subsequent study, Ledoit and Wolf (2003b) find that using a combination of

the sample covariance matrix with the constant-correlation model estimator produces even better results.

The approach that Ledoit and Wolf (2003a) have used is well-known in statistics as *shrinkage*.[6] The central idea of the original James and Stein (1961) shrinkage technique is a trade-off between estimation errors and specification errors. As the sample historical covariance matrix is asymptotically unbiased but has a lot of estimation errors, and the single-index model estimator has relatively few estimation errors but tends to have large specification errors, a successful estimator will find a compromise between the sample covariance matrix and the highly structured single-index model estimator.

The problem with the traditional covariance shrinkage estimators is that they usually break down when the number of stocks in the portfolio is large because their loss functions involve the inverse of the covariance matrix. Ledoit and Wolf (2003a, 2004) address this problem by solving the optimal shrinkage intensity from a quadratic loss function based on the Frobenius norm, which is a quadratic distance measure of the estimated covariance matrix to the observed covariance matrix and does not involve the inverse of the covariance matrix. They show that this estimator has smaller risk and is better conditioned even when the dimension of the covariance matrix is large compared to the sample size.

---

[6]The shrinkage method has been used earlier in finance but on an ad hoc basis. For example, Blume (1971) and Vasicek (1973) adjustments on the betas are also shrinkage methods.

The Ledoit and Wolf (2003a) prove that their estimator is consistent, but this is a large sample property and may not be very useful as the large T is not realistic in practice. Other researchers have published works where the simpler weighted average estimator outperforms the Ledoit and Wolf estimator.[7] For example, Jagannathan and Ma (2003) find that an equally-weighted average of the sample covariance matrix and a single-index model estimator performs as well as the more complicated Ledoit and Wolf's estimator.

Daniels and Kass (2001) provide an extensive review of the empirical Bayesian shrinkage estimators (where the empirical data determine the amount of shrinkage) proposed in recent years. They classify these methods into two general approaches. The first approach involves shrinking the unstructured estimator (the sample covariance matrix) towards a structured estimator (also called a prior or a shrinkage target) and the second approach involves shrinking the eigenvalues of the sample covariance matrix.

The distribution of a sample covariance matrix S of the covariance matrix Σ of a multivariate normal population can be described by a Wishart distribution as $S \sim W(\Sigma, n)$, where $S = \Sigma/n$ and n is the degree of freedom. The eigenvalues of the sample covariance matrix S tend to be much more dispersed than the eigenvalues of the population covariance matrix Σ, and the excess dispersion equals the error of

[7] See for example, Jagannathan and Ma (2003), Bengtsson and Holst (2002), Disatnik and Benninga (2005).

the sample covariance matrix. On average, the small eigenvalues are found to be too small and the large eigenvalues too large. Therefore an intuitively appealing approach to improve the sample covariance matrix is to shrink the sample eigenvalues towards some central value (Muirhead (1987)). A number of studies have focused on the eigenvalue estimation problem.[8] Yazici (1996) uses a simple method to adjust the eigenvalues of two successive sample covariance matrices assuming constant eigenvectors and finds that the adjusted covariance matrix not only produces more accurate covariance estimates but also leads to better estimates for portfolio risk.[9]

2.1.4 Average models

Average models are another type of structural models that are designed to improve the estimation of the sample covariance matrix. They assume that the historical data provide better information regarding the average relationships among stocks in the portfolio than the individual pair-wise ones, as there are too much sampling variations for the latter to be significant.[10] The overall mean model is the most aggregate form of the average models, which assumes that all future covariances (correlations) between stocks equal the average of the historical pair-wise covariances (correlations). It is found to outperform most alternative covariance estimation methods.[11] Unlike the

[8]See for example Stein (1977), Haff (1980, 1991) and Dey and Srinivasan (1985).

[9]A similar eigenvalue adjustment is shown in Chapter 4.

[10]The most commonly assumed average relationship is the constant correlation coefficient (see for example the studies by Elton and Gruber). However, there are also studies that assume constant covariance (see for example Chan, Karceski and Lakonishok (1999)).

[11]See for example, Elton and Gruber (1973), Elton, Gruber and Urich (1978), Chan, Karceski and Lakonishok (1999) and Elton, Gruber and Spitzer (2006).

index models, the more disaggregate average models that form groups based on some homogeneous characteristics of the stocks in the portfolio, are found to be able to improve the estimation power of the covariance matrix (Elton and Gruber (1973), Eun and Resnick (1992) and Elton, Gruber and Spitzer (2006)).

2.1.5 Time varying models

As market conditions change over time, the covariance structure of returns changes as well. A conventional method to characterize the time-varying covariance structure of returns is to employ a multivariate GARCH model.[12] The major problem with these models is that they require a large number of parameters to be estimated hence are computationally feasible for only a small number of stocks. As we are interested in estimating the return covariance matrix for a reasonably large amount of stocks, these models are not very helpful. As a result, we focus mostly on the constant covariance models. In the last essay of the thesis, we use a tractable mean reverting Ornstein-Uhlenbeck (OU) process to develop an alternative time varying model which has an economically motivated time varying covariance structure.

The RiskMetrics' exponentially weighted moving average method is also capable of capturing the dynamic feature of a covariance structure. The RiskMetrics model (1996)[13] estimates variances and covariances by assigning more weight to recent stock

[12]Kroner and Ng (1998) provide a comprehensive review of the multivariate GARCH type models. See for example, Bollerslev, Engle and Wooldridge (1988), Ng, Engle and Rothschild (1992), Engle and Kroner (1995), Ledoit, Santa-Clara and Wolf (2003).

[13]RiskMetrics is a system to assess the Value-at-Risk (VaR), or the maximum amount of potential

returns. It assumes that more distant historical information is less relevant than recent information for forecasting future stock relationships. An attractive feature of the RiskMetrics's model is that it can be written in a recursive form which can be used as a basis for making periodic updates of volatility forecasts. The key of the RiskMetrics model is to specify the optimal value of the decay factor, or the weights given to the more recent events. The smaller the decay factor, the bigger is the relative weight given to the recent events as compared to the more distant events. RiskMetrics model uses an average decay factor of 0.94 for daily volatility and 0.97 for monthly volatility.[14]

2.1.6 Random matrix theory models

More recently, Laloux et al (2000) and Plerou et al (2001) use the random matrix theory (RMT) to separate the noise and information in a sample covariance matrix. Random matrix theory (RMT) studies the asymptotic behavior of the eigenvalues and eigenvectors of random matrices when their dimensions increase without bound. It is first developed in the context of complex quantum systems where the precise nature of the interactions between subunits is unknown. Laloux, Cizeau, Potters and Bouchaud (1999) find that nearly 94% of the eigenvalues of the correlation matrix

loss due to the exposure of market risk. It uses an exponentially weighted moving average (EWMA) model to estimate variance and covariance in the process of forecasting the change of the value for a portfolio over a given time horizon.

[14]According to the RiskMetrics' published technical document in 1996, these optimal decay factors are determined from individual variance forecasts across the 450 time series that the RiskMetrics processes at that time. Further details can be found in RiskMetrics (1996) by Longerstaey and Spencer (1996).

of the S&P500 stocks returns (based on daily data during 1991-1996) agree with the RMT prediction, which suggests a considerable degree of the sample correlation matrix is random. Only a few eigenvectors (which have eigenvalues larger than the RMT-predicted upper edge eigenvalue of the random part of the correlation matrix) are found to contain information about groups of correlated firms and useful for the construction of optimal portfolios. Laloux et al. (2000), Plerou et al. (2001) and Rosenow et al. (2000) find that if prior to optimization, one filters out the lower part of the eigenvalue spectrum of the correlation matrix that contains mostly noise, then the filtered correlation matrix provides a much better prediction of future correlation than the sample correlation matrix, and the risk level of the optimized portfolios could also be improved.

2.2 Literature on comparing covariance estimators

In this section, we first review the statistical and economic evaluation criteria that have been used in the literature to compare alternative covariance estimators. We then review the specifications of the comparisons, particularly the N/T ratio used in the previous studies.

2.2.1 Comparison criteria

Elton and Gruber (1973) suggest that both statistical and economic criteria should be used to compare the accuracy of alternative covariance estimation techniques. Since

then most subsequent studies have followed this dual comparison approach.[15]

Most existing studies that compare the performance of covariance estimators have used similar evaluation methods. The statistical criteria measure the ability of different covariance estimators to estimate accurate pair-wise covariances. The popular statistical measures include root mean square error (RMSE), and other forms from the same family such as mean square error (MSE) and mean absolute error (MAE). The economic criteria measure the ability of covariance estimators in producing efficient out-of-sample portfolios. There are a few methods proposed to establish the economic significance of different covariance estimators. We review the Cohen and Pogue (1967)'s method of comparing the positions of efficient frontiers, the Elton and Gruber (1973)'s method of studying risk-return relationships over different risk levels, and the Chan, Karceski and Lakonishok (1999)'s method of comparing the performance of the minimum variance portfolios.

Statistical criteria

Mean square error (MSE) measures the average of the squared difference of the estimated and actual values. A smaller MSE indicates a smaller estimation deviation from the actual observation. Root mean square error (RMSE), which is the positive value of the square root of MSE, is an even better measure because RMSE has the

[15]For example, Elton, Gruber and Urich (1978) use the same methods to compare the performance of different beta adjustment techniques. Eun and Resnick (1992) use the mean square error (MSE) to compare the correlation estimation abilities of the multi-index models associated with the arbitrage pricing theory (APT) and the disaggregate mean models.

same unit as the measured subject, and therefore is easier to interpret.

The problem with MSE (and RMSE) measure is that it measures the average pair-wise covariance estimation errors. By breaking the covariance matrix to the element-by-element level estimates, the information contained in the structure of the covariance matrix is lost. As a result, it is purely a statistical measure and does not provide any implications regarding the use of the estimation of the covariance matrix.

Comparison over full range of risks

Cohen and Pogue (1967) compare the performances of different covariance structures by comparing the locations of the mean-variance efficient frontiers constructed according to these different covariance estimators. Subsequently, Elton and Gruber (1973) compare the risk and return relationships of portfolios on the different efficient frontiers over a number of pre-specified risk levels.

Both methods test the economic significance of a covariance matrix under the Markowitz (1952) framework. Unfortunately these methods involve too many comparisons at different risk and return levels, and do not provide a straightforward assessment of alternative covariance estimators. In addition, since they compare efficient portfolios over the full range of risks, they require information on expected returns hence their results depend on the properties of those expected returns. As

studies have found that the mean-variance optimization is very sensitive to the estimated returns[16], using expected returns effectively dilutes the importance of the covariance estimation in the portfolio selection process.

Comparison of the minimum variance portfolio

Chan, Karceski and Lakonishok (1999) use the performance of the minimum variance portfolio (MVP) to compare alternative covariance estimators. The MVP is constructed by minimizing the portfolio variance without the constraint on the target level of returns. It is the only portfolio on the efficient frontier whose weights do not depend on expected returns.[17] Comparing the performance of the MVPs therefore helps to focus on the effect of the estimation of covariances than expected returns. This method has been followed by many subsequent studies.[18]

More specifically, Chan, Karceski and Lakonishok (1999) use the optimized weights to calculate the buy-and-hold returns of the MVPs for a period of time following a

[16]See for example, Michaud (1989), Best and Grauer (1991), Chopra and Ziemba (1993), Winston (1993), Chan, Karceski and Lakonishok (1999).

[17]More specifically

$$min_w w'\Sigma w \tag{2.2.1}$$
$$s.t. w'1 = 1$$

where w_{mvp} are the weights of the minimum variance portfolio. The optimized weights of a global minimum variance portfolio equal the following

$$w_{mvp} = \frac{\Sigma^{-1}\underline{1}}{\underline{1}'\Sigma^{-1}\underline{1}} \tag{2.2.2}$$

where $\underline{1}$ is a column vector of ones, and Σ is the $N \times N$ return variance and covariance matrix.

[18]See for example, Moskowitz (2003), Jagannathan and Ma (2003) and Basak, Jagannathan and Ma (2004).

portfolio re-balancing and repeat the estimation and the optimization procedures at the end of this period for a number of years. They then use the resulting time-series of the monthly returns to characterize the performance of the optimized portfolios based on different covariance estimators.

The problem with the minimum variance portfolio is that it provides very limited information regarding the performance of covariance matrix estimators as it is based on only one special portfolio. One contribution of this thesis is to find more informative evaluation criteria that can be used to assess the risk system as a whole, in which the risk of any kind of the portfolios may need to be measured.

Many studies also impose no-short-sale restrictions and upper-bound constraints on the amount of any security in the portfolio when they study the performance of the MVPs. Jagannathan and Ma (2003) find that such constraints can improve the performance of the sample covariance matrix as in such circumstances the mean-variance optimization implicitly applies some shrinkage on the extreme sample covariance estimates toward the mean. Throughout the thesis, we do not impose these restrictions as we are interested in the maximum errors of covariance matrix estimators in predicting portfolio variances.

2.2.2 N/T ratio

Recently, Pafka and Kondor (2003, 2004) use a simulation-based approach to systematically compare the relative performance of different correlation matrix estimators

for portfolio selection and risk management. They study different values of the number of stocks N and length of the time series T and find that the ratio N/T is an important factor that influences the relative performance of alternative correlation estimation methods.

Most existing studies use empirical data to compare the performance of covariance estimators. As empirical data are often limited, many of them use monthly stock return data to estimate a covariance matrix for a much larger number of stocks. As a result, their N/T ratios are typically greater than 1.[19] This explains why many studies have found that the simpler methods such as the single-index model and the overall mean model produce better covariance estimations than the sample covariance matrix because there is a great benefit in reducing the estimation errors in the sample covariance matrix. This ratio is also important in our study when we interpret the results.

[19]For example, Cohen and Pogue (1967) use 10 annual return data to estimate the covariance matrices of 75 and 150 stocks; Elton and Gruber (1973) use 60 monthly return observations to estimate the covariance matrix of 76 stocks; Eun and Resnick (1992) use 84 monthly return data to estimate the covariance matrix of 140 stocks; Chan, Karceski and Lakonishok (1999) use 60 monthly return observations for 250 stocks; Ledoit and Wolf (2003a) use 120 monthly return data for 1000 stocks; and Jagannathan and Ma (2003) use 60 monthly return data to estimate the covariance matrix of 500 stocks.

Chapter 3

An empirical analysis of covariance estimation accuracy: old and new models

3.1 Introduction

As we have reviewed in the previous chapter, the early methods of estimating a covariance matrix can be broadly grouped into the following three categories: the sample historical covariance matrix, index models and average models. The relative performance of these different estimation methods has been well studied.[1] The general consensus of these studies are that the sample historical covariance matrix provides the worst covariance prediction; the multi-index covariance estimators provide less accurate covariance predictions than the single-index covariance estimators; and the

[1]See for example Elton and Gruber (1973), Elton, Gruber and Urich (1978), Eun and Resnick (1992), Chan, Karceski and Lakonishok (1999), and Elton, Gruber, Brown and Goetzmann (2007).

average covariance estimators provide the best covariance prediction among all the alternatives. In other words, the more complicated covariance estimators in general fail to consistently out-perform the simpler ones.

With the advancement in both optimization and computational technologies, a number of technically sophisticated covariance estimation methods have been developed. There are a few papers that study these new methods separately and find they perform better than the traditional ones.[2] However, there is no study that has compared these different new methods.

This motivates us to conduct a comprehensive examination of both old and new covariance estimation methods. An analysis of these methods together using the same dataset may provide fresh evidence regarding the relative performance of these methods. This study is intended to contribute to the literature on comparing the covariance estimation for portfolio optimization by providing an analysis of the up-to-date estimation methods.

Using empirical weekly stock return data, we study the estimation accuracy of eleven covariance estimators[3] for the returns of 78 NYSE 100 index component companies. We compare these estimators by using the most commonly used criteria in

[2]See for example, Ledoit and Wolf (2003a, 2003b), Jagannathan and Ma (2003), Pafka and Konnor (2004).

[3]These estimators include the sample historical covariance matrix, the single-index unadjusted-beta estimator, the single-index Blume-adjusted-beta estimator, the single-index Vasicek-adjusted-beta estimator, the multi-index industry factor estimator, the multi-index principal components factor estimator, and the overall mean estimator, the RiskMetrics exponentially weighted moving average method, the Ledoit and Wolf (2003a) Bayesian shrinkage method, the Jagannathan and Ma (2003) simple average method, and the random matrix filtering method.

the existing research, namely the root mean square error of pair-wise covariance estimations (RMSE) and the volatility of the minimum variance portfolios (MVP), so that our results are comparable to the findings in the existing literature.

While we have some consistent results from both the RMSE and MVP test, we find these two comparisons in general give different results and the results under the MVP differ systematically from those obtained under the RMSE. More specifically, methods based on very simple structures (such as the single-index estimators and the overall mean estimator) tend to do much worse for the MVP test than the RMSE measure. Methods imposing a richer (but still simplified) structure and adjusting for noises (such as the multi-index estimators, the Ledoit and Wolf estimator, the Jagannathan and Ma estimator and the random matrix filtering estimator), do better under the MVP test.

The RMSE is a statistical measure and does not provide any implications regarding the use of the estimated covariance matrix, while the MVP provides only limited information on the ability of the covariance estimators in predicting portfolio variances as it is based on only one special portfolio. This shows that we need to find a more powerful measure to compare the performance of alternative covariance estimators. In Chapter 4, we will propose a robust measure which is not only more powerful than the MVP as it can measure the risks of any portfolios, but also has a sound interpretation on both statistical and economic grounds and can be used as a single

measure to compare different covariance estimators. This helps to eliminate the problem of having different results when two measures are used separately.

The rest of this chapter is organized as follows: Section 3.2 reviews both the traditional and contemporary covariance estimation methods; Section 3.3 reviews the comparison criteria commonly used to assess the effectiveness of alternative covariance estimators; Section 3.4 describes the data used in this empirical study; Section 3.5 presents the empirical findings; and Section 3.6 concludes.

3.2 Description of covariance estimators

In this section we describe the eleven covariance estimation methods that we compare in the study. Seven of these methods appear earlier in the literature and have been studied more extensively.[4] The other four methods are relatively new.[5] Having reviewed most of these methods in the literature review, we concentrate on providing the details of how they are implemented.

3.2.1 Sample historical covariance matrix

The sample historical covariance matrix is obtained using the asset returns over a historical sample period. The pair-wise covariance $\sigma_{i,j}$ for stock i and j is estimated

[4]The seven old methods are the sample historical covariance matrix, the single-index covariance estimator, the single-index model with Blume-adjusted betas estimator, the single-index model with Vasicek-adjusted betas estimator, the multi-index industry factor estimator, the multi-index principal components factor estimator, and the overall mean estimator.

[5]The four new methods are the RiskMetrics exponentially weighted moving average method, the Ledoit and Wolf (2003a) Bayesian shrinkage method, the Jagannathan and Ma (2003) estimator, and the random matrix filtering method.

as

$$\hat{\sigma}_{i,j} = E[R_{i,t} - E(R_i)][R_{j,t} - E(R_j)] \tag{3.2.1}$$

where $\wedge$ indicates the estimate, $R_{i,t}$ and $R_{j,t}$ are the returns of two stocks i and j at time t, and $E(R_i)$ and $E(R_j)$ are the average return of stock i and j over a time period T where $t \in [1, T]$.

The sample historical covariance matrix is seldom used to estimate the future covariance as it is found to contain too much estimation noises, but it is often included in the comparison studies as a benchmark to evaluate the effectiveness of other covariance estimators.

3.2.2 Single-index unadjusted beta estimator

Index models are developed to explain why stocks move together and to ease the problem of large estimation parameters for the sample covariance matrix. The most well known index model is the single-index model (Sharpe (1963)), which assumes that stocks move together only because of their common responses to an aggregate market index. Specifically, we have

$$\hat{\mathbf{\Sigma}}_{single-index} = \mathbf{B}\Omega_m\mathbf{B}' + \mathbf{E} \tag{3.2.2}$$

where $\mathbf{B}$ is a $N \times 1$ vector of the stock betas β_i ($i \in [1, N]$), N is the number of stocks. Ω_m is the variance of market returns. β_i measures the sensitivity of stock's return $R_{i,t}$ to the market return $R_{m,t}$ and is estimated using ordinary least square

(OLS) regression of the sample stock returns over the market returns.[6] $\mathbf{E}$ is a $N \times N$ diagonal matrix of the idiosyncratic risks.

3.2.3 Single-index Blume-adjusted beta estimator

Blume (1971) suggests a simple linear regression adjustment to adjust the beta's regression tendency towards their grand mean. This method consists of regressing the betas from one historical period on the betas from a prior period and then using this regression to adjust the betas for the estimation period. More specifically, the unadjusted beta estimate obtained from the second period (β_2) is regressed on the unadjusted beta estimate obtained from the first period (β_1)

$$\beta_{2,i} = a + b\beta_{1,i} + c_i \tag{3.2.3}$$

where i denotes the stock in the sample. Assuming the relationship of betas between two adjacent periods is the same, the unadjusted beta estimate from the second period (β_2) is substituted into this regression equation as the independent variable to obtain an estimate of the beta for the third period β_{Blume}.

$$\beta_{Blume,i} = a + b\beta_{2,i} \tag{3.2.4}$$

[6]The single-index model assumes that

$$\begin{aligned} R_{i,t} &= \alpha_i + \beta_i R_{m,t} + e_{i,t} \\ E[e_i e_j] &= 0 (i \neq j) \\ E[e_i (R_m - \bar{R}_m)] &= 0 \end{aligned}$$

where $R_{i,t}$ and $R_{m,t}$ are returns for stock i and the market index at time t respectively, and $e_{i,t}$ is the residual return that is not explained by the stock's response to the market return. β_i is estimated using OLS and equals $\beta_i = \frac{\sigma_{im}}{\sigma_m^2} = \frac{\sum_{t=1}^{T}\left[(R_{it}-\bar{R}_{it})(R_{mt}-\bar{R}_{mt})\right]}{\sum_{t=1}^{T}\left(R_{mt}-\bar{R}_{mt}\right)^2}$.

Blume (1971) technique measures the relationship between betas over two periods and results in an extrapolation of the trend of betas in the two observed periods. Unless there is a reason to suspect such a continuous drift in betas, this is an undesirable property.

3.2.4 Single-index Vasicek-adjusted beta estimator

If we define the average beta across the sample stocks in the historical period as $\bar{\beta}$, then the Vasicek (1973) procedure involves taking a weighted average of $\bar{\beta}$ and the historical beta β_i for security i. The adjusted beta for security i equals:

$$\beta_{Vasicek,i} = \frac{\sigma^2_{\beta_i}}{\sigma^2_{\bar{\beta}} + \sigma^2_{\beta_i}}\bar{\beta} + \frac{\sigma^2_{\bar{\beta}}}{\sigma^2_{\bar{\beta}} + \sigma^2_{\beta_i}}\beta_i \tag{3.2.5}$$

where $\sigma^2_{\bar{\beta}}$ is the variance of the distribution of the historical estimates of beta over sample stocks. $\sigma^2_{\beta_i}$ is a measure of uncertainty associated with the measurement of the individual security beta (β_i) and it is the square of the standard error of the beta estimate for security i for the historical period prior to the period that the beta is estimated.[7]

Equation (3.2.5) shows that the bigger the uncertainty $\sigma^2_{\beta_i}$ about the estimate of β_i, the bigger will be the Vasicek's adjustment, as indicated by a smaller weight of $\frac{\sigma^2_{\bar{\beta}}}{\sigma^2_{\bar{\beta}}+\sigma^2_{\beta_i}}$ placed on the beta's historical estimate β_i.

---

[7] Assuming that the relationship between $R_{i,t}$ and $R_{m,t}$ is described by a stationary bivariate normal distribution, the standard error in the measurement of beta for a security is measured as $\sigma_{\beta_i} = \sigma_{ei}/\sigma_m$ (Elton, Gruber, Brown and Goetzmann (2007)).

3.2.5 Multi-index industry factor estimator

Conventional wisdom suggests that stocks belonging to the same industry group are more related because they tend to be affected by same common events. King (1966) finds that stocks have a strong industry association even after the removal of market index influences.

We review a relatively simple multi-index industry factor model where the returns of each security are assumed to be affected by the market index and an index for the industry to which the company belongs. Furthermore, each industry index is assumed to be uncorrelated with the market and with all other industry indices.[8] We have

$$R_{g,t} = a_g + b_g R_{M,t} + Res_{g,t} \tag{3.2.6}$$

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + \gamma_i Res_{g,t} + \varepsilon_{i,t} \tag{3.2.7}$$

where $R_{g,t}$ is the index return for the industry group g ($g \in [1, M]$, M is the number of industry groups). $Res_{g,t}$ is the industry residual return that is not explained by the market index and $\varepsilon_{i,t}$ is the stock residual return that is not explained by the market index and the industry index. The covariance matrix is obtained as:

$$\hat{\mathbf{\Sigma}}_{multi-industry} = \mathbf{B}\Omega_m\mathbf{B}' + \mathbf{\Upsilon}\mathbf{\Omega}_{ind}\mathbf{\Upsilon}' + \mathbf{E} \tag{3.2.8}$$

where $\mathbf{B}$ is a $N \times 1$ vector of β_i, Ω_m is the variance of market returns, $\mathbf{\Upsilon}$ is a $N \times M$

[8] To avoid the multi-collinearity problem, we run our multi-index industry model in two stages: first, the returns of different industry groups are regressed over the market index return; secondly, the individual stock returns are regressed over both the market index and the residuals from the first regression.

matrix of industry factor loadings, and $\mathbf{\Omega}_{ind}$ is the variance matrix of the industry indices.[9] $\mathbf{E}$ is a $N \times N$ diagonal matrix of idiosyncratic risks.

3.2.6 Multi-index principal components estimator

The principal components are extracted from the sample data (Jackson (1991)). Compared to other types of multi-factor models that use macroeconomic or stock fundamental factors, the advantage of using principal components (PC) is that these PC are orthogonal to each other and ranked according to their ability to best explain the historical correlation matrix. We have

$$R_{it} = \alpha_i + b1_i R_{pc1t} + b2_i R_{pc2t} + \ldots + bp_i R_{pcpt} + e_{it} \tag{3.2.9}$$

where $R_{pc1t}, R_{pc2t}, ...R_{pcpt}$ are the return indices of the principal components and $b1, b2, ...bp$ are the corresponding factor loadings. p is the number of principal components included in the model. The covariance matrix of the above principal components model is calculated as:

$$\hat{\mathbf{\Sigma}}_{multi-PC} = \mathbf{B}\mathbf{\Omega}_{PC}\mathbf{B}' + \mathbf{E} \tag{3.2.10}$$

where $\mathbf{B}$ is a $N \times p$ matrix with each column representing the loadings of stocks on a principal component factor, and $\mathbf{\Omega}_{PC}$ is a $p \times p$ diagonal matrix with the diagonal elements equal the variance of each principal components. $\mathbf{E}$ is a $N \times N$ diagonal matrix of idiosyncratic risks.

---

[9] Since we assume the industry indices are orthogonal to each other, Ω_{ind} contains only diagonal elements that equal the variance of the industry indices.

Connor and Korajczyk (1993) find between one to six factors for the NYSE and AMEX stock returns over 1967-1991. We choose to use five factors for our study.[10]

3.2.7 Overall mean estimator

The overall mean model is first introduced by Elton and Gruber (1973). It predicts the future stock correlations all equal the average of historical pair-wise correlations, that is

$$\hat{\mathbf{\Sigma}}_{correlation} = \begin{bmatrix} 1 & & \bar{\rho}_{ij} \\ & \ddots & \\ \bar{\rho}_{ij} & & 1 \end{bmatrix} \tag{3.2.11}$$

where $\bar{\rho}_{i,j}$ is equal to the average historical pair-wise correlations $\bar{\rho}_{i,j} = \frac{2}{n(n-1)} \sum(\rho_{i,j})$.

Elton and Gruber (1973) find that the overall mean model, despite of its over simplifying nature, performs surprisingly well compared to the sample historical correlation matrix and single index correlation estimator.

3.2.8 RiskMetrics estimator

The RiskMetrics method adjusts return series so that the latest observations carry the highest weight in the volatility estimation, so

$$\mathbf{R} = \sqrt{\frac{1-\lambda}{1-\lambda^T}} \begin{pmatrix} r_t^{(1)} & r_t^{(2)} & \cdots & r_t^{(N)} \\ \sqrt{\lambda} r_{t-1}^{(1)} & \cdots & \cdots & \sqrt{\lambda} r_{t-1}^{(N)} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\lambda^{T-1}} r_{t-(T-1)}^{(1)} & \sqrt{\lambda^{T-1}} r_{t-(T-1)}^{(2)} & \cdots & \sqrt{\lambda^{T-1}} r_{t-T}^{(N)} \end{pmatrix} \tag{3.2.12}$$

[10]We also use a simple scree test to justify the choice of five PC factors for our sample data in Appendix B.

where $\lambda \in (0, 1]$ is the optimal decay factor, r_t denotes the return from day $(t-1)$ to t, and $\mathbf{R}$ is a $T \times N$ matrix of adjusted returns. T is the length of time-series, and N is the number of stocks in the portfolio. The covariance between stock i and j is calculated as:

$$\hat{\sigma}_{i,j}^2 = \sum_{t=1}^{T} \frac{\lambda^{t-1}}{\sum_{t=1}^{T} \lambda^{t-1}} (r_{i,t} - \bar{r}_i)(r_{j,t} - \bar{r}_j) = \mathbf{R}^T \mathbf{R} \tag{3.2.13}$$

An attractive feature of this estimator is that it can be written in a recursive form, that is

$$\sigma_{ij,t+1|t}^2 = \lambda \sigma_{ij,t|t-1}^2 + (1 - \lambda) r_{i,t} r_{j,t} \tag{3.2.14}$$

Equation (3.2.14) is convenient to use to periodically update the covariance forecasts. Based on its published technical document in 1996 (Longerstaey and Spencer (1996)), RiskMetrics model uses an average decay factor of 0.94 for daily volatility and 0.97 for monthly volatility.

3.2.9 Ledoit and Wolf Bayesian shrinkage estimator

Ledoit and Wolf (2003a) propose to take a weighted average of the historical covariance matrix and the single-index covariance estimator to obtain an optimal trade-off of the estimation errors and specification errors in the two estimators. That is,

$$\hat{\mathbf{\Sigma}}_{shrinkage} = \hat{\alpha}^* \hat{\mathbf{\Sigma}}_{single-index} + (1 - \hat{\alpha}^*) \hat{\mathbf{\Sigma}}_{historical} \tag{3.2.15}$$

where $\hat{\alpha}^*$ is the estimated optimal weight for combining the single-index covariance estimator and the sample historical covariance matrix. Ledoit and Wolf (2003a)

derive $\hat{\alpha}^*$ by solving a quadratic loss function based on the Frobenius norm. or $||\alpha\hat{\mathbf{\Sigma}}_{single-index} + (1-\alpha)\hat{\mathbf{\Sigma}}_{historical} - \mathbf{C}||^2 = \sum\sum(\alpha s_{ij} + (1-\alpha)h_{ij} - \sigma_{ij})^2$, where s_{ij}, h_{ij} and σ_{ij} are the entries of the single-index covariance matrix, the sample historical covariance matrix and the observed covariance matrix respectively.[11]

3.2.10 Jagannathan and Ma simple average estimator

Jagannathan and Ma (2003) estimator takes a simple average of the sample historical covariance matrix and the single-index covariance estimator. That is,

$$\hat{\mathbf{\Sigma}}_{J\&M} = \frac{1}{2}\hat{\mathbf{\Sigma}}_{single-index} + \frac{1}{2}\hat{\mathbf{\Sigma}}_{historical} \tag{3.2.16}$$

[11]More specifically,

$$\begin{aligned}
\hat{\alpha}^* &= max[0, min[\frac{\hat{\kappa}}{T}, 1]] \\
\hat{\kappa} &= \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}} \\
\hat{\pi} &= \sum_{i=1}^{N}\sum_{j=1}^{N} AsyVar[\sqrt{T}h_{ij}] \\
\hat{\rho} &= \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} AsyCov[\sqrt{T}s_{ij}, \sqrt{T}h_{ij}] \\
\hat{\gamma} &= \sum_{i=1}^{N}\sum_{j=1}^{N}(h_{ij} - s_{ij})^2
\end{aligned}$$

where $\hat{\pi}$ denotes the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$, $\hat{\rho}$ denotes the sum of the asymptotic covariances of the entries of the single-index covariance matrix with the entries of the sample covariance matrix scaled by $\sqrt{T}$, and $\hat{\gamma}$ measures the mis-specification of the (population) shrinkage target. Further details can be found in Ledoit and Wolf (2003a, 2003b, 2004).

3.2.11 Random matrix filtering estimator

For a $T \times N$ return matrix (where N is the number of stocks and T is the length of time series), when $N \to \infty$, $T \to \infty$ and $Q = T/N \geq 1$, the random matrix theory (RMT) predicts the density of the eigenvalues of the correlation matrix $\mathbf{C}_{corr}$, $\rho_C(\lambda)$, equals (Laloux, Cizeau and Potters (2000))

$$\rho_C(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{3.2.17}$$

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \tag{3.2.18}$$

where $\lambda_\pm$ are the minimum and maximum eigenvalues of the correlation matrix and $\lambda_+ > \lambda_i > \lambda_-$.

Laloux et al. (2000) suggest to improve the correlation matrix by filtering out the noises in the matrix by adjusting its eigenvalues. The eigenvalue matrix of the historical correlation matrix, $\mathbf{\Lambda}$,[12] is filtered to contain only eigenvalues that are larger than the maximum eigenvalue λ_+ predicted by RMT. The eigenvalues that are smaller than the maximum eigenvalue λ_+ are replaced by an average of all the smaller eigenvalues to preserve the trace of the correlation matrix. The idea is that since the eigen-states corresponding to the *noise band* are not expected to contain real information, one should not distinguish the different eigenvalues and eigenvectors

[12]Using eigen-decomposition, we can obtain the eigenvalues Λ and eigenvectors T of the historical covariance matrix $\Sigma_{historical}$, as $\Sigma_{historical} = T\Lambda T'$.

in this sector. So

$$\mathbf{\Lambda}_{filtered} = \begin{bmatrix} \lambda_n & & & & 0 \\ & \ddots & & & \\ & & \lambda_k & & \\ & & & \bar{\lambda} & \\ & & & & \ddots \\ 0 & & & & \bar{\lambda} \end{bmatrix} \tag{3.2.19}$$

where $\lambda_n > \ldots > \lambda_k > \lambda_+$. $\bar{\lambda}$ is the average of residuals so that the trace of the matrix is preserved, i.e., $\mathrm{Tr}(\mathbf{\Lambda})$=$\mathrm{Tr}(\mathbf{\Lambda}_{filtered})$.

The RMT filtered correlation matrix is obtained as:

$$\hat{\mathbf{\Sigma}}_{RMT-filtered} = \mathbf{T}\mathbf{\Lambda}_{filtered}\mathbf{T}' \tag{3.2.20}$$

where T is the eigenvector of the historical covariance matrix $\hat{\mathbf{\Sigma}}_{historical}$.

3.3 Description of performance measures

In this section, we describe the two comparison criteria we use to compare the eleven covariance estimators. These two criteria, namely the root mean square error (RMSE) of the pair-wise covariance estimations and the risk analysis of the minimum variance portfolios (MVP), have been used most often in the related literature.

3.3.1 Root mean square error

We use the root mean square error to compare the pair-wise estimation accuracy of covariances. The root mean square error (RMSE) has the same unit as the measured

subject therefore is easier to interpret. A smaller RMSE indicates a smaller estimation deviation from the actual observation.

RMSE is defined as

$$RMSE = \sqrt{\frac{N(N-1)}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} (\hat{\sigma}_{i,j} - \sigma_{i,j})^2} \tag{3.3.1}$$

where $N(N-1)/2$ is the number of the pair-wise covariances for a $N \times N$ covariance matrix; $\hat{\sigma}_{i,j}$ and $\sigma_{i,j}$ are the pair-wise estimated and the actual covariances respectively.

3.3.2 MVP test

We follow Chan, Karceski and Laknoishok (1999) to examine the volatility of the minimum variance portfolios constructed according to the covariance estimators. There are however two differences in the way we conduct the analysis. Firstly, we form a global minimum variance portfolio using each estimated covariance matrix, without imposing any restrictions on portfolio weights. Using the weights of the minimum variance portfolio, we calculate and record the out-of-sample returns of the minimum variance portfolios. The resulting time-series of the returns then allow us to evaluate the volatility of the minimum variance portfolios. Secondly, unlike Chan, Karceski and Laknoishok (1999) which re-balance portfolios every year to get the return time-series, we assume the portfolio holding period is the same as that of the return observations. That is, we assume the minimum variance portfolios are

buy-and-hold for the entire 3 years of the out-of-sample period.

3.4 Data

3.4.1 Stock returns

We use Datastream's Wednesday to Wednesday weekly dividend-adjusted stock prices for the NYSE US 100 index constituent companies (as of 11 May 2006) from 06 January 1988 to 25 December 1996. To be consistent with most studies, we try to avoid periods with extreme events happening in the market because the data is usually distorted during these periods. Our sample data avoids the 1987 major market crash and 1998 financial crisis.

We use the NYSE US 100 index stocks because these are large stocks and are more likely to be included in actual portfolios. These stocks represent a large percentage of the market, hence together they form a portfolio that is closer to the market portfolio than using the same number of other stocks. 78 out of the 100 index constituent stocks have the full history for the period of our study. The detailed descriptions of these stocks are included in Appendix A.[13]

[13]Note that 22 of the 100 component stocks are excluded from the sample because they do not have the complete return data for the entire period that we study. We compared the summary statistics of these 22 stocks and the stocks included in the sample for a period of 10.11.2004 to 09.11.2005 when all stocks have complete return data (not reported). We find that on average the risks of the excluded stocks are not very different from the stocks of the same industries that are included in our sample. Although we select the stocks from the index components constructed later than the period we study, this selection bias affects more on the average returns rather than the covariances that we are concerned with. Therefore, excluding these 22 stocks should not affect the overall conclusions of our study.

Most traditional covariance studies usually use monthly data (Elton, Gruber. Brown and Goetzmann (2007)). Using monthly data makes our time-serie observations fairly small when compared to the size of our portfolio, which may result in a lot of noise in the sample covariance matrix. Using daily or intra-day higher frequency data increases the number of time-serie observations, but recent studies find that the cost related to confounding microstructure issues can be large. Weekly data therefore seems to be a good compromise and has been used in a number of recent studies (Moskowitz (2003), Elton, Gruber and Spitzer (2006)).

The reason to use Wednesday to Wednesday closing prices is due to the existing evidence of high autocorrelations using Friday to Friday and low autocorrelations using Monday to Monday prices (Chordia and Swaminathan (2000)). Wednesday seems a natural compromise (again) and it is also used in Moskowitz (2003).

To be consistent with the existing literature, we calculate stock returns as the natural logarithm of prices:

$$R_{weekly} = ln(P_{t+1}/P_t) \tag{3.4.1}$$

Throughout the thesis, we use returns in weekly and percentage terms, i.e., 0.0003 means 0.0003% per week. We can easily transform these weekly returns and volatilities into annualized terms if needed as

$$R_{annualized} = R_{weekly} \times 52 \tag{3.4.2}$$

$$SD(R)_{annualized} = SD(R)_{weekly} \times \sqrt{52} \tag{3.4.3}$$

3.4.2 Indices returns

We use the S&P 500 as the proxy for the market index. S&P 500 also has industry return indices for the ten industry sectors to classify stocks, namely Basic materials, Consumer (Cyclical), Consumer (Non-Cyclical), Energy, Financial, Healthcare, Industrial, Technology, Telecommunications and Utilities. Unfortunately, the S&P 500 industry index data is available only since January 1995, which is after our sample period starts.

We therefore construct our own industry indices using the S&P 500 industry classification information. Our industry indices equal the average return of the sample stocks that belong to the same industry groups. That is

$$R_{Gi,t} = \frac{1}{n}\sum_{i=1}^{n} R_{i,t} \tag{3.4.4}$$

where stock i is one of the n stocks that belong to the industry group Gi, R_i is the stock i's return and R_{Gi} is the industry index return for the industry group Gi.

3.4.3 Sub-sample period

Following most studies in this area (Elton, Gruber and Urich (1978), Chan, Karceski and Lakonishok (1999)), we divide the nine years into three three-year sub-periods: 01.1988-12.1990, 01.1991-12.1993 and 01.1994-12.1996. The first sub-period return data is used to generate the initial covariance estimation; the information from the

second sub-period is then used to adjust this initial estimation; and the third sub-period is used as the out-of-sample to evaluate the accuracy of the estimation. For methods that do not require a second-period adjustment, this process is reduced into two: the second sub-period is used to generate the estimation, which is then compared to the out-of-sample. Our N/T ratio is less than 0.5 (78/157), which is quite low compared to most related studies which usually have a N/T ratio of at least 1.

Most studies (for example Elton, Gruber and Urich (1978)) use a five-year sub-period. We choose a three-year sub-period because: first, unlike previous studies that use monthly returns, we use weekly return data. Hence a three-year sub-period gives us enough observations to generate a reasonably good sample covariance matrix. Secondly, in our study there are no restrictions on portfolio construction, which makes our results more pertinent to active portfolio management. Since the holding period by an active manager is on average shorter than that of a typical long-term fund manager, three-year seems to be reasonable.

3.5 Empirical Results

3.5.1 Descriptive statistics

To set the stage for our analysis, we first calculate the basic descriptive statistics for the sample stocks used in the study. Table 3.1 reports the average sample weekly

Table 3.1: **Sample return, volatility, covariance and correlation** This table reports the summary statistics of the 78 sample stock returns for the three sub-sample period of 01.1988-12.1990, 01.1991-12.1993 and 01.1994-12.1996, as well as for the entire period of study of 01.1988-12.1996. The returns are reported in percentage terms, for example, 0.0030 means 0.0030%. The first sub-sample is used to form the estimation, the second sub-sample is used to adjust the initial estimation and the third sub-sample is used as the out-of-sample realization to evaluate the estimation performance.

Sample period	Mean	Stdev	Covariance	Correlation
1st sub-sample 1988-1990	0.2359	3.7114	4.6420	0.3545
2nd sub-sample 1991-1993	0.3332	3.7078	3.4136	0.2465
3rd sub-sample 1994-1996	0.3418	3.2110	2.2380	0.2251
Combined 1988-1996	0.3037	3.5672	3.4181	0.2747

stock returns, variances, covariances and correlations of the three sub-periods as well as the entire sample period of our study.

We find that all three sub-periods are not significantly different from each other and that of the entire combined period. However, we find that over the latter part of the three sub-periods, the stocks tend to have higher returns, lower volatilities and lower correlations with other stocks. As a result, if we use the prior sub-sample to predict the later sub-sample, we may find that the predicted variances and covariances are above the out-of-sample values.

We report the summary statistics of the variables employed in the covariance estimators in Appendix B.

Table 3.2: **Average estimated variances and covariances** This table reports the average and the cross-sectional standard deviation of the estimated variances and pair-wise covariances of eleven covariance estimators. The predictions are made for the period of 01.1994 to 12.1996, using the weekly returns information from 01.1988 to 12.1993. The variances and covariance of the realized weekly returns from 01.1994 to 12.1996 are also reported as the out-of-sample values for comparison purposes. Average estimated variance and covariance give some indication of the overall predicted pair-wise relationships.

	Average variance		Average covariance	
	mean	std	mean	std
out-of-sample	11.0950	7.2088	2.2380	1.6550
Historical	14.9680	10.0830	3.4136	2.4388
Single Unadjusted	14.9680	10.0830	2.5598	1.3960
Single Blume	15.5680	10.1750	3.1810	1.4977
Single Vasicek	14.5820	9.0179	2.5377	0.0694
Multi Industry	14.9680	10.0830	2.9803	2.1787
Multi PC	14.9680	10.0830	3.4843	2.5241
Overall Mean	14.9680	10.0830	3.3844	1.4415
RiskMetrics	14.7910	9.8961	3.2995	2.3844
Ledoit & Wolf	14.9680	10.0830	3.3954	2.0546
Jagannathan & Ma	14.9680	10.0830	2.9867	1.8159
Random Matrix	14.6880	9.0187	3.4147	2.2632

3.5.2 Average estimated covariance

Table 3.2 reports the average of the estimated variances and pair-wise covariances and the corresponding cross-sectional standard deviations for the different estimators. We can see that the average estimated variances and covariances are bigger than those of the out-of-sample values.

For the average of the variances, there are only small differences among the estimations of the different estimators except for the single-index Blume-adjusted-beta estimator, which has the biggest average variance estimate. When we check the estimated betas (reported in Appendix B), we find that the estimated betas are on average bigger in the second sub-period than those in the first sub-period and the out-of-sample period. The Blume (1971) adjustment method picks up the extrapolation trend of the estimated betas from the first sub-period to the second sub-period, where in fact there is no such an increasing trend. As a result, the single-index Blume estimator predicts higher betas and variances.

In terms of the average estimated covariance, we find that the single-index unadjusted beta estimator and the single-index Vasicek adjusted beta estimator have closer average covariance predictions to the out-of-sample values than the other estimators.

Although the above results give some information regarding the variance and covariance predictions by the different estimators, the average estimate value is a very rough measure because the over-estimation and under-estimation cancel out in the averaging process. A better measure of the estimation errors is the RMSE, which penalizes errors of estimations in both directions.

Table 3.3: **RMSE results** This table reports the root mean square error of the pair-wise estimated covariances obtained from the eleven covariance estimators with the corresponding out-of-sample values. The covariance predictions are based on weekly returns from 01.1988 to 12.1993, and compared with the out-of-sample covariances from 01.1994 to 12.1996. The estimation estimators are sorted in order of the worst to the best performance.

	RMSE
Multi PC	2.4184
Historical	2.3105
RiskMetrics	2.2193
Random Matrix	2.2086
Ledoit & Wolf	2.0686
Overall Mean	2.0501
Multi Industry	1.9004
Single Blume	1.8802
Jagannathan & Ma	1.7719
Single Vasicek	1.6563
Single Unadjusted	1.6143

3.5.3 RMSE

Table 3.3 reports the RMSE of the estimated pair-wise covariances against their out-of-sample realized values.[14] The estimation methods are sorted in the order of the worst to the best in terms of RMSE performance.

We find that the sample historical covariance matrix performs rather badly and under-performs almost all other covariance estimators. In addition, we find that in general the simpler estimators perform better than the more complicated estimators.

[14]Although not reported here, we find that the results of the RMSE for the estimated correlations agree mostly although not entirely with those for the covariances. This indicates that there are some slight differences on the relative performance of the estimators depending on whether the estimation is made for covariance or correlation.

For example, the top four methods include all three single-index estimators. and the Jagannathan and Ma estimator out-performs the more complicated Ledoit and Wolf estimator.

We also find that the multi-index industry estimator performs better than the multi-index PC estimator. This shows that general industry factors are better than the sample-specific principal components factors in predicting pair-wise covariances.

We suspect that the under-performance we find for the Ledoit and Wolf estimator and the random matrix filtering estimator may be related to the low ratio of number of stocks to the time series observations (N/T) in our sample. In Ledoit and Wolf (2003a), their N/T ratio is over 8, whereas in our study the ratio is less than 0.5. In addition, Pafka and Kondor (2003) find that noise has the effect suggested by the random matrix theory (RMT) for relatively large values of N/T. In other words, the random matrix filtering estimator may be more useful when we have a larger portfolio with few sample observations.

3.5.4 MVP test

Table 3.4 reports the means and standard deviations of the minimum variance portfolios (MVPs) constructed using different covariance estimators. We are interested in the standard deviations of these portfolios, which represent the risk an investor is exposed to over the out-of-sample period. The average returns are also reported but they are for descriptive purposes only.

Table 3.4: **MVP test results** This table reports the average weekly performance of the minimum variance portfolios that are constructed based on alternative covariance estimators. Given the optimal portfolio weights constructed using the information from 01.1988 to 12.1993, we record the weekly returns of the minimum variance portfolios in the out-of-sample 3-year periods from 01.1994 to 12.1996 and calculate the average mean and volatility of the time-series of returns. The covariance estimators are ranked in the order of the highest to the lowest volatility of the MVP.

	mean	std
Overall Mean	0.0060	1.8096
Historical	0.1739	1.6649
RiskMetrics	0.1718	1.6368
Single Blume	0.0527	1.5160
Single Unadjusted	0.0849	1.4342
Multi PC	0.0881	1.4062
Ledoit & Wolf	0.1426	1.4059
Single Vasicek	0.2098	1.3375
Random Matrix	0.1503	1.3193
Multi Industry	0.2185	1.3179
Jagannathan & Ma	0.1405	1.2738

We find that the relative performances of different estimators based on the MVP test differ systematically from those obtained under the RMSE criterion. More specifically, the performances of the simpler estimators on average decrease, while the performances of more complicated methods improve dramatically. For example, two single-index estimators (the single-index Blume estimator and the single-index unadjusted estimator) that are among the top performers under the RMSE perform rather badly under the MVP test. On the other hand, the three best estimators under MVP

(the Jagannathan and Ma estimator, the multi-index industry estimator and the random matrix filtering estimator), are all new methods that impose a richer structure in different ways. This shows that the best diversified portfolios require additional information that is not available in a single-index model.

There are some findings consistent with those of the RMSE. For example, the sample historical covariance matrix under-performs most of the other models; the Jagannathan and Ma estimator out-performs the Ledoit and Wolf estimator; and the multi-index industry estimator out-performs the sample specific multi-index PC estimator.

3.5.5 Summary

To summarize, although we have some similar results from the RMSE and MVP test, these two comparisons in general give different results as they measure different things. Overall, we find that methods based on very simple structures (such as the single-index estimators and the overall mean estimator) tend to do much worse under the MVP test than the RMSE measure. Methods imposing a richer (but still simplified) structure and adjusting for noises (such as the multi-index estimators, the Ledoit and Wolf estimator, the Jagannathan and Ma estimator and the random matrix filtering estimator), do better under the MVP test.

More importantly, the difference of our results from the previous studies may be related to our relatively low N/T ratio compared to those in the other studies. Given

that we are using empirical data, we cannot examine the sampling variation of these results. In order to better assess the performances of the estimators, we will need a simulation method to help us better understand their sampling properties.

3.6 Conclusion

In this chapter, we review both old and new covariance estimation methods and empirically compare the performance of eleven covariance estimators using two comparison criteria commonly used in the related literature, namely the RMSE and the MVP test. We find that the methods imposing a richer structure and adjusting for noises do better under the MVP test. However, given that the MVP only measures the performance of one single portfolio, we want a more robust assessment where any portfolios can be measured. In addition, we want to explore the sampling variations of the results in a controlled way. This leads to our next chapter, where we propose a more robust portfolio-based eigen distance measure to compare two estimated covariance matrices.

Chapter 4

Risk model appraisal: a new portfolio-based eigen-distance measure

4.1 Introduction

It is well known that the ranking of covariance estimators should depend on both the quality of the estimators and the measurement criteria (Makridakis et al. (1982, 1993, 2000)). Depending on the use of a covariance matrix and the purpose of the comparison, the differences of covariance matrix estimators may be different. Using comparison criteria that do not fit the purpose of the comparison of covariance matrix estimators may produce very misleading results. Therefore, as part of exploring the best way to estimate covariance matrices. it is necessary to first consider the best way to appraise such estimates.

Furthermore, as we have discussed earlier, the minimum variance portfolio test provides very limited information regarding the ability of a covariance estimator in predicting portfolio variances as it only measures the performance of one special portfolio. As a covariance matrix is increasingly used in portfolio risk management, we need a new criterion that can be used to measure the risks of any kind of portfolios. This chapter is motivated by the need to search for such a suitable appraisal criterion.

In this chapter, we propose a new portfolio distance measure based on eigen decomposition (*eigen-distance*) to compare different covariance estimators in terms of how far the portfolio risks estimated by two covariance matrices can differ from each other. We prove that this new measure is a proper measure of distance. We show that this innovative criterion is robust and particularly suitable for applying to a risk measurement system as a whole, where any portfolio may need to be considered. It is operational and can be easily applied to evaluate empirical covariance matrices of large portfolios. It can also be applied equally well to both the absolute variances of portfolios and the variances of their tracking errors against a benchmark.

We use simulations to study the sampling properties of the proposed criterion. To run our simulation, we have to choose a *true* covariance structure to simulate from. We construct this *true* covariance matrix by taking a historical covariance matrix and adjusting its eigenvalues to reduce the sampling variation (which we discuss later in Section 4.3.1). This is a well-known statistical technique to improve the

sample covariance matrix, which we will discuss in more details in Chapter 6. This kind of adjusted covariance matrix is found to have most empirical properties of financial returns and at the same time contains considerably less amount of noises than the covariance matrices estimated using other methods (Yazici (1996)). We simulate the stock returns by decomposing this true covariance matrix using a N-variate normal model. These N factors are defined to be orthogonal and multivariate normally distributed with zero means and variances equal the eigenvalues of the true covariance matrix. The factor loadings are defined as the corresponding eigenvectors. We then use the simulated return data to construct the estimated covariance matrices according to different covariance matrix estimators and compare these matrices with both the *true* covariance matrix and an ex-post sample from the *true* covariance matrix. As there are only sampling errors in the simulation world and the *true* covariance matrix structure is exactly known, we can study precisely the performance of this criterion in assessing various covariance estimators and the significance of these results. Our objective is to examine whether the comparison results obtained using our proposed criterion are consistent.

The rest of this chapter is organized as the following. In Section 4.2 we derive our new portfolio-based eigen-distance measure. We first introduce the notion of an eigen-portfolio and show how the eigenvector and eigenvalue are related to a conventional portfolio and portfolio return variance. We then show how our portfolio-based

eigen-distance is designed to measure the biggest differences between the portfolio risk predictions under two covariance matrices. We also present the statistical and economic interpretations of this distance measure. Section 4.3 describes the simulation experiment and our covariance decomposition method. Section 4.4 reports the simulation results, which we use to study the sampling properties of the new eigen-distance criterion. We also compare these properties with those of the RMSE criteria. Section 4.5 concludes.

4.2 A new portfolio-based eigen-distance measure

Suppose there are N assets in a portfolio and we want to compare two covariance matrices $\mathbf{V_1}$ and $\mathbf{V_2}$.[1] In this section we introduce a new approach in the form of a distance measure, $d(\mathbf{V_1}, \mathbf{V_2})$, to compare the performance of two covariance matrices on both statistical and economic grounds.

We first make two important assumptions regarding the covariance structure and the portfolios of interest.

Assumption 4.2.1. *Both covariance matrices are non-singular (strictly positive definite).*

[1] Our proposed eigen-distance can be used to measure the difference of any two covariance matrices. We illustrate the comparison between two estimated covariance matrices. But the comparison could also be between a estimated covariance matrix and a true or ex-post realized covariance matrix.

This assumption allows us to do eigen decomposition without any problem and obtain a finite distance measure. It also makes sure that all eigenvalues of the covariance matrices are strictly positive.

Assumption 4.2.2. *There is no restriction on portfolio weights, i.e., they are not constrained to be non-negative or bounded.*

The no restriction on weights assumption simplifies the derivation of analytical results. More importantly, it is necessary for a system to adequately measure the tracking error variance relative to a benchmark. If we use $\mathbf{x}$ and $\mathbf{b}$ to represent the portfolio and benchmark weights respectively, the difference of the portfolio weights from the benchmark equals $\mathbf{z} = \mathbf{x} - \mathbf{b}$ ($\sum z_i = 0$) and the variance of the portfolio tracking error (TE) equals $Var(TE) = \mathbf{z}'\mathbf{V}\mathbf{z}$ (suppose $\mathbf{V}$ is the return covariance matrix of stocks in the portfolio). In order for the sum of the weights differences $\sum z_i$ to be zero, there must be short positions in the portfolio unless $x = b$. This shows that we need to have short positions in the portfolio to measure the tracking error. We need to allow short positions in order to calculate the variance of the portfolio tracking error.

4.2.1 Eigen-portfolios and eigenvalue variances

As our eigen-distance measure uses the eigenvectors and eigenvalues of a covariance matrix, we first establish a link between them and traditionally defined portfolios and

portfolio return variances. We show that an eigen-portfolio, which is a portfolio with weights equal to an eigenvector, has its return variance equal to the corresponding eigenvalue.

The words *eigenvalues* and *eigenvectors* are derived from the word *eigen* in German, which means *characteristic*. In geometric terms, eigenvectors define the orthogonal directions in space and eigenvalues define the scaling on these orthogonal directions. A symmetric and non-singular matrix can be completely characterized by a unique set of non-zero eigenvectors and their corresponding eigenvalues.

The eigen decomposition theorem decomposes a matrix into matrices composed of its eigenvectors and eigenvalues. For any symmetric matrix $\mathbf{V}$ (i.e. any covariance matrix), we can apply the eigen decomposition theorem and obtain the following:

$$\mathbf{V} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}' \tag{4.2.1}$$

where $\mathbf{\Lambda}$ is a $N \times N$ eigenvalue matrix with eigenvalues λ_n ($n \in [1, N]$) on the diagonal. $\mathbf{T}$ is a $N \times N$ eigenvector matrix with each column representing a corresponding eigenvector $\mathbf{t}_n$, where $\mathbf{T} = [\mathbf{t_1}, \mathbf{t_2}, \ldots, \mathbf{t_N}]$ and $\mathbf{T}'\mathbf{T} = \mathbf{I}$. $\mathbf{I}$ is an identity matrix.

Equation (4.2.1) can be rearranged and written as the following:

$$\mathbf{T}'\mathbf{V}\mathbf{T} = \mathbf{\Lambda} \tag{4.2.2}$$

If we define an eigen-portfolio as the portfolio with weights equal to an eigenvector $\mathbf{t}_n$, Equation (4.2.2) shows that this eigen-portfolio has return variance equal to its

corresponding eigenvalue λ_n and is independent of any other eigen-portfolios $\mathbf{t}_m$ ($m \in [1, N], m \neq n$).

The only problem we seem to have here is the interpretation of eigenvalues as measures of portfolio risk. Since the eigenvectors do not sum up to 1 ($\sum t_i^2 = 1$ but $\sum t_i \neq 1$), we cannot interpret the eigenvalues as the variance of the percentage returns obtained by holding a portfolio.[2] In principle, we could always convert the eigenvalue to a risk measure (standard deviation (SD)) by simply applying the following

$$\text{SD of \% Return} = \sqrt{\frac{\lambda}{(\sum t_i)^2}} \tag{4.2.3}$$

In practice, though, we will instead just consider directly the variance (or SD) of the dollar returns on portfolios normalized so that their squared holdings sum to 1.

We show in the following that this normalization works very well for interpreting the first eigenvalue as the variance of the market portfolio. The first eigen-portfolio can be regarded as close to the market portfolio (Brown(1989)). As its loadings are relatively equally weighted and $\sum t_i^2 = 1$, this gives an average loading of $t_i = \frac{1}{\sqrt{N}}$ and the summation of the loadings as $\sum t_i = \sqrt{N}$. If we normalize λ to be λ/N, then $\sum t_i^2 = \frac{1}{N}$, the average loading equals $t_i = \frac{1}{N}$ and the summation of the loadings equals $\sum t_i = 1$. For example, as we show in Table 4.1 later, λ for the first eigen-portfolio equals 405.4 per week. Given that we have $N = 78$ stocks, this translates

[2] If we measure t_i in dollar units, the variance is in the square of dollar units, not in the square of percentages that are used more often.

to a reasonable annual standard deviation of 16.4% ($= \sqrt{\frac{405.4}{78} \times 52}$) for the market factor. However, for more general eigen-portfolios, this interpretation becomes less useful as the loadings on stocks are no longer equally weighted.

4.2.2 Eigen-distance comparison of two covariance matrices

Definition 4.2.1. *We define the portfolio-based eigen-distance,* $d(\mathbf{V_1}, \mathbf{V_2})$*, as the natural logarithm of the maximum ratio of the variances of any two portfolios,* x *and* y*, under the two covariance matrices* $\mathbf{V_1}$ *and* $\mathbf{V_2}$*:*

$$d(\mathbf{V_1}, \mathbf{V_2}) = \log\left(\frac{\max_{\mathbf{x}} \frac{\mathbf{x'V_1x}}{\mathbf{x'V_2x}}}{\min_{\mathbf{y}} \frac{\mathbf{y'V_1y}}{\mathbf{y'V_2y}}}\right) \tag{4.2.4}$$

where x *and* y *define the two extreme eigen-portfolios which are predicted to have the most different return variances under two covariance matrices* $\mathbf{V_1}$ *and* $\mathbf{V_2}$*, such that* x *has more risk under covariance matrix* $\mathbf{V_1}$ *(than* $\mathbf{V_2}$*); and* y *has more risk under* $\mathbf{V_2}$ *(than* $\mathbf{V_1}$*).*[3]

Definition 4.2.2. *We also define a new covariance matrix* $\mathbf{V}$*, which will be useful in our analysis, as*[4]

$$\mathbf{V} = \left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)' \mathbf{V_2} \left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right) \tag{4.2.5}$$

where $\mathbf{T_1}$ *and* $\mathbf{\Lambda_1}$ *are the eigenvector matrix and eigenvalue matrix of the covariance matrix* $\mathbf{V_1}$ *respectively. This gives the covariance matrix under* $\mathbf{V_2}$ *of portfolios that*

[3] We use small letters x and y to denote the portfolios, and bold letters $\mathbf{x}$ and $\mathbf{y}$ to denote the weights of these two portfolios.

[4] Throughout the derivation, we demonstrate the eigen decomposition on the covariance matrix $\mathbf{V_1}$. The results are the same if we do the eigen decomposition on the covariance matrix $\mathbf{V_2}$. We show in the proof that which way to start the decomposition does not affect the eigen-distance.

are independent and with unit variance under $\mathbf{V}_1$. *We will call this the* $\mathbf{V}_1$ to $\mathbf{V}_2$ distortion matrix, *as it is* $\mathbf{I}$ *when* $\mathbf{V}_1 = \mathbf{V}_2$.

Theorem 4.2.1. *The eigen-distance,* $d(\mathbf{V_1}, \mathbf{V_2})$, *can be calculated as:*

$$d(\mathbf{V_1}, \mathbf{V_2}) = \log\left(\frac{\lambda_{max}}{\lambda_{min}}\right) \tag{4.2.6}$$

where λ_{max} *and* λ_{min} *are the largest and smallest eigenvalues of our distortion matrix* $\mathbf{V}$. *The larger the* $d(\mathbf{V_1}, \mathbf{V_2})$, *the larger is the difference between the two covariance matrices* $\mathbf{V_1}$ *and* $\mathbf{V_2}$, *and vice versa.*

Proof of Theorem 4.2.1.

If we do a complete eigen-decomposition on the covariance matrix $\mathbf{V_1}$, we have

$$\mathbf{V_1} = \mathbf{T_1}\mathbf{\Lambda_1}\mathbf{T_1'} \tag{4.2.7}$$

$$\left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)' \mathbf{V_1} \left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right) = \mathbf{I} \tag{4.2.8}$$

where $\mathbf{I}$ is a $N \times N$ identity matrix.

Equation (4.2.8) shows that if we carefully adjust the coordinate system of the the covariance matrix $\mathbf{V_1}$, we can have the same unit length ($\mathbf{I}$) on every orthogonal direction defined by $\left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)$. This also means that eigen-portfolios defined by $\left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)$ are expected to be orthogonal and have unit return variances under the estimated covariance matrix $\mathbf{V_1}$. As we want to find the largest prediction difference under two covariance matrices, these eigen-portfolios form a natural basis for examining the second covariance matrix $\mathbf{V_2}$. We can find out how these portfolios perform

under $\mathbf{V_2}$, and we want to find the two portfolios for which the predicted variances are as different as possible under the two covariance matrices.

Equation (4.2.5) defines the new covariance matrix $\mathbf{V}$ under $\mathbf{V_2}$ of the scaled eigen-portfolios which are orthogonal and have unit variance under $\mathbf{V_1}$. To find the portfolio with the largest possible variance under $\mathbf{V_2}$ that has unit variance under $\mathbf{V_1}$, we would solve the following problem:

$$\max_x \mathbf{x}'\mathbf{V}\mathbf{x} \qquad \text{s.t.} \qquad \mathbf{x}'\mathbf{x} = \mathbf{1}$$

which is equivalent to finding the eigenvalues and eigenvectors of $\mathbf{V}$, since the first order condition of the above problem is simply $(\mathbf{V} - \lambda\mathbf{I})x = 0$.

If we do the eigen-decomposition on $\mathbf{V}$ and find its largest and smallest eigenvalues of λ_{max} and λ_{min}, then these two eigenvalues represent the most extreme portfolio variances under $\mathbf{V_2}$ relative to those under $\mathbf{V_1}$, as indicated in Equation (4.2.6).

Finally, we prove that our portfolio-based eigen-distance $d(\mathbf{V_1}, \mathbf{V_2})$ satisfies all four required conditions as a proper distance function. It is important to note that if one matrix is a scalar multiple of the other so $\mathbf{V}_1 = k\mathbf{V}_2$, then $d(\mathbf{V}_1, \mathbf{V}_2) = 0$. To strictly satisfy the requirements of a distance norm we should therefore only regard $\mathbf{V}_1$ and $\mathbf{V}_2$ as equivalent if no such scalar exists. It is a proper distance norm for correlation matrices, or matrices which are normalized in any way, for example, scaled to provide the same variance for all equally weighted portfolios. The detailed proof is shown in Appendix C.

This completes the proof of Theorem 4.2.1.

Corollary 4.2.1. *In terms of the original assets, the two eigen-portfolios x and y that have the most different return variances under two covariance matrices $\mathbf{V_1}$ and $\mathbf{V_2}$, have weights equal to the following:*

$$\mathbf{x} = \left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)\mathbf{t}_{max} \tag{4.2.9}$$

$$\mathbf{y} = \left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)\mathbf{t}_{min} \tag{4.2.10}$$

x is predicted to have unit variance under $\mathbf{V_1}$ but the largest variance λ_{max} under $\mathbf{V_2}$; and y is predicted to have unit variance under $\mathbf{V_1}$ but the smallest variance λ_{min} under $\mathbf{V_2}$. $\mathbf{t}_{max}$ and $\mathbf{t}_{min}$ are the corresponding eigenvectors of the largest eigenvalue (λ_{max}) and the smallest eigenvalue (λ_{min}) of $\mathbf{V}$.

Corollary 4.2.1. provides a useful guide to identify two extreme portfolios that have the most different return variances under two covariance matrices.

4.2.3 Interpretation

We know that in geometry, eigenvectors define the orthogonal directions in space and eigenvalues measure the length on these orthogonal directions. A complete eigen decomposition on covariance matrix $\mathbf{V_1}$, as shown by Equation (4.2.8), indicates that in the eigen-space defined by the adjusted eigenvectors $\left(\mathbf{T_1}\mathbf{\Lambda_1}^{-\frac{1}{2}}\right)$, the covariance matrix $\mathbf{V_1}$ has unit length in all orthogonal directions. $\mathbf{V}$ can be viewed as a transformed covariance matrix $\mathbf{V_2}$ under the same adjusted eigen space. where $\mathbf{V_1}$ has

the unit length for all coordinates. Under this new eigen space, $\mathbf{V_2}$ has the largest and smallest lengths, λ_{max} and λ_{min}, of all orthogonal directions and $\mathbf{x}$ and $\mathbf{y}$ show the corresponding coordinates. Therefore, in addition to measure the biggest differences of one covariance matrix from the other in terms of the most different portfolio variance predictions under the two covariance matrices, our proposed eigen-distance $d(\mathbf{V_1}, \mathbf{V_2})$ can also be viewed as a simple measure of the geometric difference of two matrices in terms of their most different lengths of all directions under the same coordinate system.

Having explained its interpretation, our proposed eigen-distance $d(\mathbf{V_1}, \mathbf{V_2})$ can be used as a single measure that has a sound statistical and economic basis to compare two covariance matrices. This will eliminate the problem of having different results when the statistical and economic tests are used separately.

We have mentioned earlier that our eigen-distance can be used to measure the difference of any two covariance matrices. If we measure the difference between one estimated covariance matrix and the actual ex-post covariance matrix, then the larger the eigen-distance, the more different is the estimated covariance matrix from the actual covariance matrix. Naturally, if we compare the eigen-distances of different estimated covariance matrices to the actual covariance matrix, then the estimated covariance matrices with smaller eigen-distances are closer to the actual covariance matrix, and vice versa.

In summary, the eigen-distance measure that we have introduced has the following desirable properties. It provides more information than does the MVP test. It can assess the risk systems as a whole where any portfolio (not only any absolute portfolios but also any tracking error portfolios) can be measured. It has the nice interpretation of the maximum error possible in measuring portfolio risks. Using it as a single measure, we can eliminate the potential conflicting results when the two measures (i.e., RMSE and MVP) are conducted separately. It therefore provides more useful information than these most commonly used comparison criteria in the existing literature. Finally, it is operational and can be easily applied to portfolios with a reasonably large number of stocks.

This eigen-distance measure also has its limitations. As we have shown that since the weights of the portfolio do not sum up to 1 (but the square of the weights sum up to 1), we cannot interpret the portfolio variance as the variance of the percentage returns obtained by holding a portfolio. In addition, technically the eigen-distance is a proper distance measure for the correlation matrices and normalized covariance matrices. It regards two covariance matrices as close even though they may differ substantially in scale. Theoretically, it could be problematic but empirically, this would not be much of a problem since almost all covariance estimators give similar estimations.

4.3 Simulations

To examine the performance of our proposed portfolio-based eigen-distance, we need to study its sampling properties. We employ a simulation method to do so as it appears to be difficult to obtain the results we are interested in analytically. Our objective is to examine whether the results obtained using various samples from a dataset of known statistical properties are consistent.

It is important to choose appropriate parameters so that the simulation results can represent as many properties of the empirical data as possible. On the other hand, the historical covariance matrix contains errors that need to be adjusted for. Therefore our desired simulation world is based on historical returns but at the same time makes variance adjustments for the effects of sampling variation. We use an adjusted empirical covariance matrix as the *true* covariance matrix. This adjusted empirical covariance matrix is constructed by adjusting the eigenvalues of two covariance matrices assuming eigenvectors are stable.

We then simulate artificial returns by decomposing this *true* covariance matrix using a N-variate factor model and use the simulated return data to construct estimated covariance matrices according to different covariance estimators. We compare these estimated covariance matrices with both the *true* covariance matrix and an *ex-post* sample of the *true* covariance matrix and analyze the comparison results.

In the remaining of this section, we first describe how we obtain our assumed *true*

covariance matrix. We then show how to decompose this *true* covariance matrix to generate the artificial return data.

4.3.1 Specifying a *true* covariance structure

To run our simulation, we have to choose a *true* covariance matrix to simulate from. We construct this *true* covariance matrix by taking a historical covariance matrix and adjusting its eigenvalues to reduce the sampling variation. Yazici (1996) finds this kind of adjusted covariance matrix has most empirical properties of financial returns and at the same time contains considerably less amount of noises than the covariance matrices estimated using other methods.

The empirical data used to construct the historical sample covariance matrix include two three-year successive sub-sample periods of 01.1988-12.1990 and 01.1991-12.1993, for 78 NYSE 100 Composite Index component stocks. We first work out the sample covariance matrices for the two sub-sample periods as $\mathbf{C_1}$ and $\mathbf{C_2}$. We decompose $\mathbf{C_1}$ to find its eigenvalues $\mathbf{\Lambda_1}$ and eigenvectors $\mathbf{T_1}$:

$$\mathbf{C_1} = \mathbf{T_1 \Lambda_1 T_1'} \tag{4.3.1}$$

Assuming the eigenvectors are stable,[5] we adjust the eigenvalue matrix $\mathbf{\Lambda_1}$ over the second sub-sample period as the following:

$$\mathbf{T_1' C_2 T_1} = \mathbf{\Lambda} \tag{4.3.2}$$

[5]We have checked the Pearson Rank Correlation of the two eigenvalue series from the adjacent covariance matrices is 99.46%, which indicates that the principal components are remarkably stable, i.e., we can assume that the eigenvectors are stable.

This process can also be regarded as calculating the variance of the eigen-portfolios in the second sub-sample period. As eigen-portfolios are orthogonal to each other, $\mathbf{\Lambda}$ should be a diagonal matrix.[6] We therefore construct a new matrix $\mathbf{\Lambda}^*$ by keeping the diagonals and discarding the off-diagonals of $\mathbf{\Lambda}$. We call $\mathbf{\Lambda}^*$ the adjusted eigenvalue matrix and it has the following form:

$$\mathbf{\Lambda}^* = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Finally, the adjusted covariance matrix $\mathbf{C_{true}}$, which we regard as the *true* covariance matrix, is obtained as:

$$\mathbf{C_{true}} = \mathbf{T_1 \Lambda^* T_1'} \tag{4.3.3}$$

The method we use here belongs to a family of well-known statistical techniques to improve the estimation of a sample covariance matrix called the *eigenvalue shrinkage method*. We will discuss and develop this kind of covariance estimator in more detail in Chapter 6.

Table 4.1 reports the ten largest and ten smallest unadjusted and adjusted eigenvalues, the percentage return variances they explain, and the percentage adjustment both in terms of the eigenvalues and their explained return variances. We can see that unadjusted eigenvalues have a much larger dispersion than the adjusted eigenvalues.

[6] We have tested the off-diagonal values of $\mathbf{\Lambda}$ using the Fisher Z transformation and t-test and found they are all statistically insignificant from zeros. Hence these off-diagonals represent noises and discarding them should not affect the properties of $\mathbf{\Lambda}$.

Table 4.1: **Summary of unadjusted and adjusted eigenvalues** This table presents the unadjusted and adjusted eigenvalues, the percentage of return variances explained by the eigenvalues and the percentage adjustment of the adjusted over unadjusted eigenvalues. Panel A presents the results of the 10 largest eigenvalues and Panel B reports those of the 10 smallest eigenvalues. The sample periods used are from 01.1988 to 12.1990 and from 01.1991 to 12.1993.

	Eigenvalues			%variance explained		
rank	unadjusted	adjusted	%adjustment	unadjusted	adjusted	%adjustment
Panel A. largest 10 eigenvalues						
1	405.4100	297.3300	-26.6600	34.4220	25.4660	-26.0170
2	96.8630	64.2630	-33.6560	8.2242	5.5041	-33.0740
3	56.9780	48.4580	-14.9520	4.8377	4.1505	-14.2070
4	44.6890	34.5400	-22.7100	3.7943	2.9584	-22.0330
5	37.6400	29.3040	-22.1470	3.1958	2.5099	-21.4640
6	33.3040	21.8340	-34.4410	2.8277	1.8701	-33.8660
7	31.0160	38.4170	23.8610	2.6334	3.2904	24.9470
8	28.6600	15.4400	-46.1280	2.4334	1.3224	-45.6550
9	25.6810	16.8300	-34.4660	2.1805	1.4415	-33.8910
10	23.7990	21.4250	-9.9749	2.0207	1.8351	-9.1854
Panel B. smallest 10 eigenvalues						
69	0.9521	6.7048	604.2000	0.0808	0.5743	610.3700
70	0.9066	4.5512	401.9800	0.0770	0.3898	406.3800
71	0.8223	6.4967	690.1100	0.0698	0.5564	697.0400
72	0.7884	4.1293	423.7600	0.0669	0.3537	428.3500
73	0.6446	4.7490	636.7900	0.0547	0.4068	643.2500
74	0.6081	5.0082	723.6200	0.0516	0.4290	730.8400
75	0.5580	3.8979	598.5400	0.0474	0.3339	604.6700
76	0.5030	4.6845	831.2400	0.0427	0.4012	839.4000
77	0.4414	4.3779	891.7600	0.0375	0.3750	900.4600
78	0.3794	6.1770	1528.0000	0.0322	0.5291	1542.2000

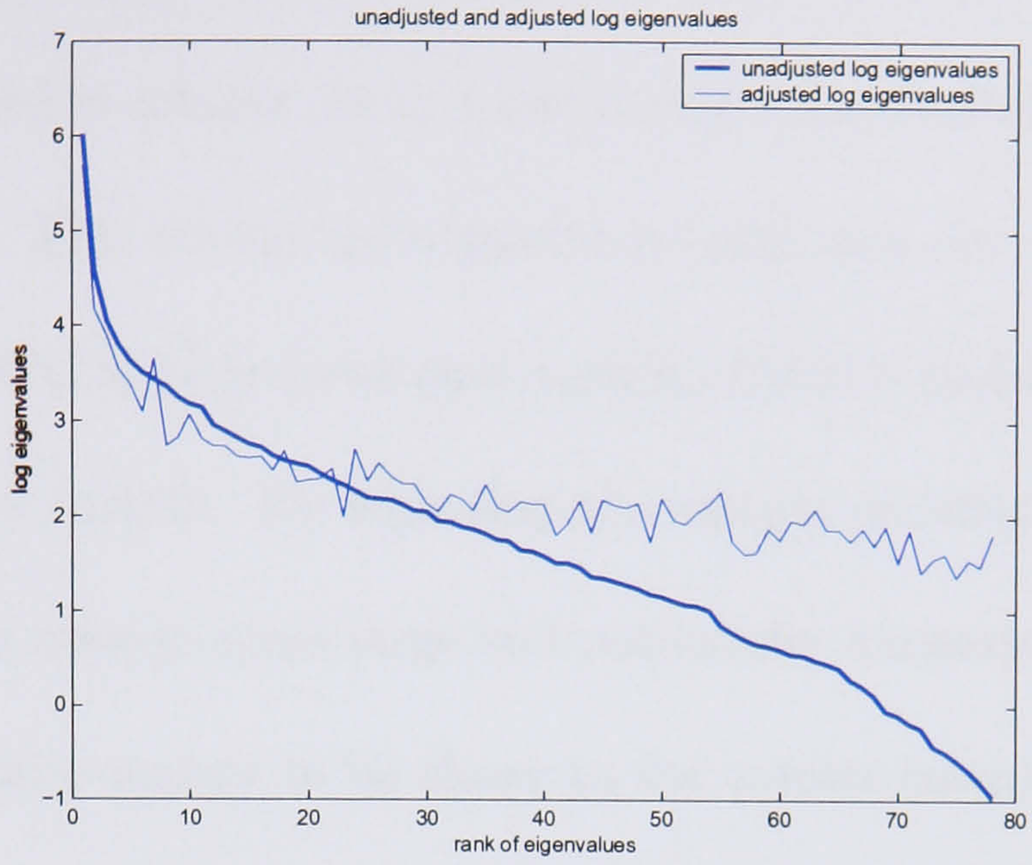


Figure 4.1: **Unadjusted and adjusted log eigenvalues** This figure plots the log values of the unadjusted and corresponding adjusted eigenvalues. The unadjusted eigenvalues are sorted from the largest to the smallest. The horizontal axis is the rank of the eigenvalues and the vertical axis is the log value of the eigenvalues.

The adjustment basically pulls the large eigenvalues and small eigenvalues towards each other. If we expect the population of all eigenvalues to have the same distribution, then the adjustment shows that bigger eigenvalues have positive measurement errors and smaller eigenvalues have negative measurement errors. Figure 4.1 plots the 78 unadjusted and adjusted log eigenvalues and visually confirms the effect of the adjustment.

The third column in Table 4.1 (percentage adjustment in terms of the change in the eigenvalues) also shows that smaller eigenvalues are adjusted upwards by bigger amounts than those of the larger eigenvalues that are adjusted downwards. The average downwards adjustment for the ten largest eigenvalues is about 20%, while the

average upwards adjustment for the ten smallest eigenvalues is over 700%. Similar results are also observed in column six in Table 4.1 for the percentage adjustment of the explained variances. This amount of adjustments indicates the size of sampling variation in the unadjusted sample covariance matrix, which is undesirable for predicting the future covariance matrix. By adjusting the sample covariance matrices through their eigenvalues, we remove some large and potentially dangerous biases. We expect the adjusted covariance matrix to be closer to the correct covariance structure.

4.3.2 Simulating returns from a *true* covariance matrix

To simulate returns that satisfy the properties of the assumed *true* covariance matrix $\mathbf{C_{true}}$, we need to find the return matrix[7] $\mathbf{R}$ such that $\mathbf{R'R} = \mathbf{C_{true}}$. It is important to emphasize that the choice of $\mathbf{R}$ is not unique. There are different methods to decompose $\mathbf{C_{true}}$ that will result in different $\mathbf{R}$'s[8]. We use a simple N-variate factor model to decompose the covariance matrix. These N factors are defined to be orthogonal and multivariate normally distributed with zero expected returns and variances equal to the eigenvalues of the *true* covariance matrix $\mathbf{C_{true}}$. The loadings for the

[7] $\mathbf{R}$ can also be regarded as a factor loading matrix, if we generate returns as $\mathbf{r} = \mathbf{R'u}$, where $\mathbf{u} \sim N(0, 1)$. In this case, $E[\mathbf{rr'}] = \mathbf{R'}E[\mathbf{uu'}]\mathbf{R} = \mathbf{R'IR} = \mathbf{R'R}$.

[8] Two most popular methods for decomposing a covariance matrix are the Cholesky decomposition and the singular value decomposition (SVD). One important difference between the two is that the Cholesky algorithm fails to provide a decomposition when the covariance matrix is not positive definite, while we can always find the SVD of a matrix. A non-positive definite covariance matrix corresponds to a situation where at least one of the risk factors is redundant, meaning that we can reproduce the redundant risk factor as a linear combination of other risk factors.

factors are defined as the corresponding eigenvectors. So for any k^{th} factor,

$$f_k \propto \mathbf{N}(0, \lambda_k^*), k = 1, \cdots, N \tag{4.3.4}$$

where λ_k^* is the k^{th} eigenvalue of $\mathbf{C_{true}}$. The loading for the k^{th} factor on each stock i, $b_{k,i}$, equals

$$b_{k,i} = t_{i,k} \tag{4.3.5}$$

where $t_{i,k}$ corresponds to the element of the i^{th} row and k^{th} column of the eigenvector matrix. The return of stock i, r_i, equals

$$r_i = \sum_{k=1}^{N} t_{i,k} f_k \tag{4.3.6}$$

Since we have assumed that the covariance matrix is non-singular, all its eigenvalues and eigenvectors are distinct and linearly independent. Our factorizing method makes sure that all factors are orthogonal to each other and all returns are explained by the N factors. The factor structure also makes it easier to separate the sources of returns.

We simulate T time series observations for each stock, so we have a $T \times N$ return matrix[9] $\mathbf{R}$, which equals

$$\mathbf{R} = (\mathbf{T_1 F})' \tag{4.3.7}$$

where $\mathbf{F}$ is a $N \times T$ matrix containing the time series of returns for the N factors, and

[9]The return matrix $\mathbf{R}$ is formed to be consistent with its previous definition as a $T \times N$ matrix.

Table 4.2: **Descriptive statistics of empirical sample and simulation returns** This table reports the mean, standard deviation, average covariance and correlation of the two empirical sample periods that we use to construct our *true* covariance matrix as well as a snapshot of the average of a 1,000 simulations of three different time-series horizons with 200, 600 and 1,000 weeks observations respectively.

	Mean	Stdev	Covariance	Correlation
Sub-sample1	0.2359	3.7114	4.6420	0.3545
Sub-sample2	0.3332	3.7078	3.4136	0.2465
Entire sample	0.2847	3.7204	4.0146	0.2972
Simulation T200	-0.0133	3.7833	3.3172	0.2312
Simulation T600	-0.0122	3.7897	3.3574	0.2338
Simulation T1000	-0.0074	3.7917	3.3526	0.2334

$\mathbf{T_1}$ is a $N \times N$ eigenvector matrix containing the factor loadings for the N factors:

$$\mathbf{F} = \begin{bmatrix} \mathbf{f_1} \\ \mathbf{f_2} \\ \vdots \\ \mathbf{f_N} \end{bmatrix} = \begin{bmatrix} f_{1,1} & \cdots & f_{1,T} \\ f_{2,1} & \cdots & f_{2,T} \\ \vdots & \cdots & \vdots \\ f_{N,1} & \cdots & f_{N,T} \end{bmatrix}; \mathbf{T_1} = [\mathbf{t_1}, \mathbf{t_2}, \ldots, \mathbf{t_N}] = \begin{bmatrix} t_{1,1} & \cdots & t_{1,N} \\ t_{2,1} & \cdots & t_{2,N} \\ \vdots & \cdots & \vdots \\ t_{N,1} & \cdots & t_{N,N} \end{bmatrix}$$

We can verify that the simulated returns have the same expected covariance structure as that of the *true* covariance matrix $\mathbf{C_{true}}$, as

$$cov(\mathbf{R}) = \mathbf{R'R} = \mathbf{T_1 F F' T_1'} = \mathbf{T_1 \Lambda^* T_1'} = \mathbf{C_{true}} \tag{4.3.8}$$

As for the market return, we simulate it as an equally weighted average of the simulated returns of the stocks in the sample.

Table 4.2 reports the summary statistics of our simulated data as well as those of the empirical samples that we use to construct our *true* covariance matrix. We can see that the standard deviations, covariances and correlations of the simulated

returns over different lengths of time series are consistent and close to those of the empirical samples. The average simulated returns are however different from the empirical samples and are close to zeros. This is because we assume the returns are generated from the factors with zero mean returns.

4.4 Sampling properties of the new measure

In this section, we present the simulation evidence of the performance of our proposed portfolio-based eigen-distance. We compare it with the RMSE measure in terms of its ability to select a superior covariance estimator.[10] We use three covariance estimators, namely the sample historical covariance matrix, the single-index unadjusted covariance estimator and the single-index Blume adjusted betas covariance estimator, to provide a simple but powerful illustration on the relative performance of our eigen-distance measure over the RMSE. These three estimators are selected because our empirical studies in Chapter 3 find that the sample historical covariance matrix performs badly under both the RMSE and MVP test, while the single-index estimators perform much worse under the MVP test than the RMSE.

[10] As our eigen-distance measure is designed to be more powerful than the MVP test, we basically replace the MVP test with the eigen-distance measure and compare its results with the RMSE.

Table 4.3: ***True* versus average estimated covariances** This table presents the average and standard deviation of the simulated covariances of three covariance estimators over 100 times simulations, when the lengths of the estimation window (T) are 200, 600 and 1,000 weeks respectively. The results of those of the *true* covariance matrix and the *ex-post* covariance matrix are also reported. The difference between the *true* covariance matrix and the *ex-post* covariance matrix is that the *true* covariance matrix does not contain any noise while the *ex-post* covariance matrix contains noises due to the finite-length of the time-series.

	T=200		T=600		T=1,000	
	Average	Stdev	Average	Stdev	Average	Stdev
true	3.3580	1.6333	3.3580	1.6333	3.3580	1.6333
ex-post	3.3172	1.9211	3.3574	1.7303	3.3526	1.6921
Historical (H)	3.2864	1.9071	3.3833	1.7439	3.3734	1.7020
Single-index (S)	3.4309	1.5355	3.5284	1.4783	3.5184	1.4595
Single-Blume (B)	3.4320	1.2950	3.5288	1.3884	3.5186	1.4062

4.4.1 Simulation results

Table 4.3 reports the mean and the standard deviation of the pair-wise covariances[11] (average of 100 simulations) of the three covariance estimators and an *ex-post* sample of the *true* covariance matrix, as well as those of the *true* covariance matrix.

We find that the difference between the average covariances estimated using the single-index estimator and the single-index Blume estimator is not as big as we find in the empirical studies in Chapter 3. This shows that the Blume adjustment works better in the simulation experiment as there is no spurious trend in the estimated betas for the Blume method to pick up during the beta adjustment process.

[11] We do not compare the variance estimation performance because all three covariance estimators produce the same estimated variances. The results of the estimated correlation coefficients are similar to those of the estimated covariances, therefore for the brevity of reporting, they are not reported either.

Table 4.4: **Eigen-distance, RMSE and their differences** This table reports one snapshot of the average eigen-distance and RMSE results of the estimated covariance matrices obtained from three covariance estimators when the estimation windows are set to 200, 600 and 1,000 weeks. The reported average results and associated standard errors are based on 100 simulations. The pair-wise differences of the eigen-distance and RMSE results of the three estimators are also reported, along with the corresponding t-statistics. $*$ indicates that the result is statistically significant at a 95% confidence level.

	T=200				T=600				T=1,000			
	vs. *true*	se	out-of-sample	se	vs. *true*	se	out-of-sample	se	vs. *true*	se	out-of-sample	se
Panel A. Eigen-distance												
Historical (H)	2.8355	0.0095	4.1608	0.0122	1.4573	0.0039	2.0736	0.0049	1.1097	0.0029	1.5751	0.0042
Single-index (S)	2.2907	0.0028	3.6823	0.0111	2.2574	0.0017	2.6633	0.0059	2.2503	0.0014	2.4826	0.0056
Single-Blume (B)	2.2911	0.0028	3.6833	0.0115	2.2580	0.0018	2.6643	0.0059	2.2503	0.0013	2.4833	0.0056
H-S	0.5448	0.0098	0.4785	0.0149	-0.8001	0.0042	-0.5897	0.0074	-1.1407	0.0031	-0.9075	0.0072
	(55.5*)		(32.2*)		(-189.8*)		(-79.6*)		(-363.9*)		(-126.5*)	
H-B	0.5444	0.0098	0.4775	0.0150	-0.8007	0.0042	-0.5907	0.0074	-1.1406	0.0031	-0.9082	0.0072
	(55.6*)		(31.8*)		(-189.7*)		(-80.2*)		(-365.5*)		(-126.2*)	
S-B	-0.0004	0.0013	-0.0010	0.0015	-0.0007	0.0004	-0.0011	0.0004	0.0000	0.0002	-0.0007	0.0003
	(-0.31)		(-0.66)		(-1.83)		(-2.40*)		(0.10)		(-2.50*)	
Panel B. RMSE												
Historical (H)	1.1006	0.0097	1.5408	0.0118	0.6335	0.0059	0.8807	0.0076	0.4917	0.0043	0.6924	0.0053
Single-index (S)	1.0999	0.0104	1.5448	0.0122	0.9242	0.0059	1.1050	0.0079	0.8801	0.0041	1.0053	0.0059
Single-Blume (B)	1.0863	0.0090	1.5321	0.0122	0.9292	0.0057	1.1093	0.0077	0.8835	0.0038	1.0083	0.0056
H-S	0.0007	0.0052	-0.0004	0.0060	-0.2906	0.0034	-0.2244	0.0038	-0.3884	0.0032	-0.3129	0.0041
	(0.13)		(-0.67)		(-85.4*)		(-58.8*)		(-119.6*)		(-76.0*)	
H-B	0.0143	0.0039	0.0087	0.0048	-0.2957	0.0033	-0.2287	0.0039	-0.3918	0.0032	-0.3159	0.0041
	(3.67*)		(1.81)		(-89.2*)		(-58.1*)		(-121.4*)		(-77.6*)	
S-B	0.0136	0.0037	0.0127	0.0035	-0.0050	0.0016	-0.0043	0.0016	-0.0034	0.0010	-0.0030	0.0011
	(3.72*)		(3.67*)		(-3.12*)		(-2.62*)		(-3.40*)		(-2.77*)	

Table 4.4 reports the eigen-distance and RMSE results of the three covariance estimators in terms of the comparison with both the *true* covariance matrix and an *ex-post* sample of the *true* covariance matrix, averaged over 100 simulations. Also reported are the pair-wise differences of the eigen-distance and RMSE results for the three estimators. The t-statistics of these differences are reported and the significance is indicated at a 95 percent confidence level.

We can see that our eigen-distance is a consistent and powerful measure to differentiate the sample covariance matrix and the single index estimators, even in small samples. For example, the difference of the sample covariance matrix and the single index unadjusted estimator (H-S) has a t-statistic of 55.5 when they are compared against the true covariance matrix with T=200 weeks.

The eigen-distance is also more powerful than the RMSE. The t-statistics of the eigen-distance differences are always more significant than those of the corresponding RMSE differences. In addition, while the eigen-distance differences can differentiate two estimators with statistical significance, the RMSE differences cannot. For example, the t-statistic of the RMSE difference of (H-S) is 0.13 when compared against the true covariance matrix and T=200 weeks, while we have discussed earlier that the corresponding t-statistic for the eigen-distance difference is 55.5.

Given that the amount of most empirical financial observations is often small (and the possibility of a structural change also makes the useful data for estimation

shorter even when we have a large amount of time series data), the power of the eigen-distance to choose a relatively superior covariance estimator with small amount of data is particularly useful.

However, we also find that the eigen-distance measure is not very well suited to differentiate two covariance estimators that are relatively close together. In this case, the single index estimator and the single index Blume estimator are not found to be statistically different. In other words, we need to find a more powerful measure to differentiate these covariance estimators. We study such a measure in Chapter 5.

In terms of the sampling properties of the relative performance of the different covariance estimators, we find the performance of the sample historical covariance matrix increases with the time series observations more than the other two single-index estimators, under both the RMSE and eigen-distance measures. For example, the eigen-distances against the true covariance matrix for the historical estimator decrease from 2.8355 to 1.1097 when T increases from 200 to 1000, while the corresponding values for the single-index estimator (B) decrease from 2.2907 to 2.2503. In addition, when T=200, the differences of both the RMSE results and eigen-distance results show that the single-index estimator performs better than the sample historical covariance matrix (for example under RMSE (H-B)=0.0007). But the opposite is true when the time series observations increase to 600 weeks and 1.000 weeks ((H-B)=-0.2906 and -0.3884). This shows that when the covariance structure is constant,

the more data we have (or the smaller the N/T ratio, as N is fixed to 78 in this case), the better the sample historical covariance matrix estimates the covariance matrix. However, when the data is limited (or N/T is large), there are benefits to use simpler single-index estimators.

Finally, we find that although the absolute values of both the eigen-distance and RMSE of the estimators are bigger when they are compared to the ex-post covariance matrix than when they are compared to the *true* covariance matrix, their pair-wise differences are smaller. For example, the eigen-distances of H with the *true* covariance matrix and with the ex-post covariance matrix when T=200 equal 2.8355 and 4.1608 respectively; and those of S equal 2.2907 and 3.6823. The corresponding differences between H and S however, equal 0.5448 and 0.4785. This indicates that due to the sampling variation in the ex-post covariance matrix, it makes the estimation more difficult to all estimators, which leads to relatively smaller differences among their performances.

4.4.2 Robustness check with increasing lengths of estimation window

In this section, we check the robustness of the performance of the eigen-distance and RMSE with different lengths of the estimation window. We plot in Figure 4.2 the pair-wise differences of the eigen-distance and RMSE of the three covariance estimators when the time series observations in the sample increase from 100 weeks to 1,000

weeks, and we examine the comparisons with both the *true* covariance matrix and the *ex-post* covariance matrix.

We can see from Figure 4.2 that the patterns of the pair-wise differences of the eigen-distance results are similar to those of the RMSE results. All three pair-wise differences are positive when the time series observations are small and negative when the time series observations are large. The differences decrease monotonically, indicating that the performance of the historical covariance matrix improves more than the single-index estimators do when the length of the estimation window increases.

We find that while the patterns of the results obtained under the RMSE and eigen-distance measure are similar, the cross over points where the relative performance of the sample covariance matrix and the single index estimators reverses are different. For the eigen-distance measure, this happens when T=300 weeks. While for the RMSE, it happens around T=200 weeks.

We also find that the differences of the RMSE and eigen-distance between the single-index estimator and the single-index Blume estimator are close to zero, and do not vary much with different lengths of the estimation window.

In addition, we find that the pair-wise differences of both RMSE and eigen-distance are smaller when the RMSE and eigen-distance are measured against the ex-post covariance matrix than against the *true* covariance matrix itself. This is shown by the smaller ranges of plots of the Figure B and D as compared to the Figure A and

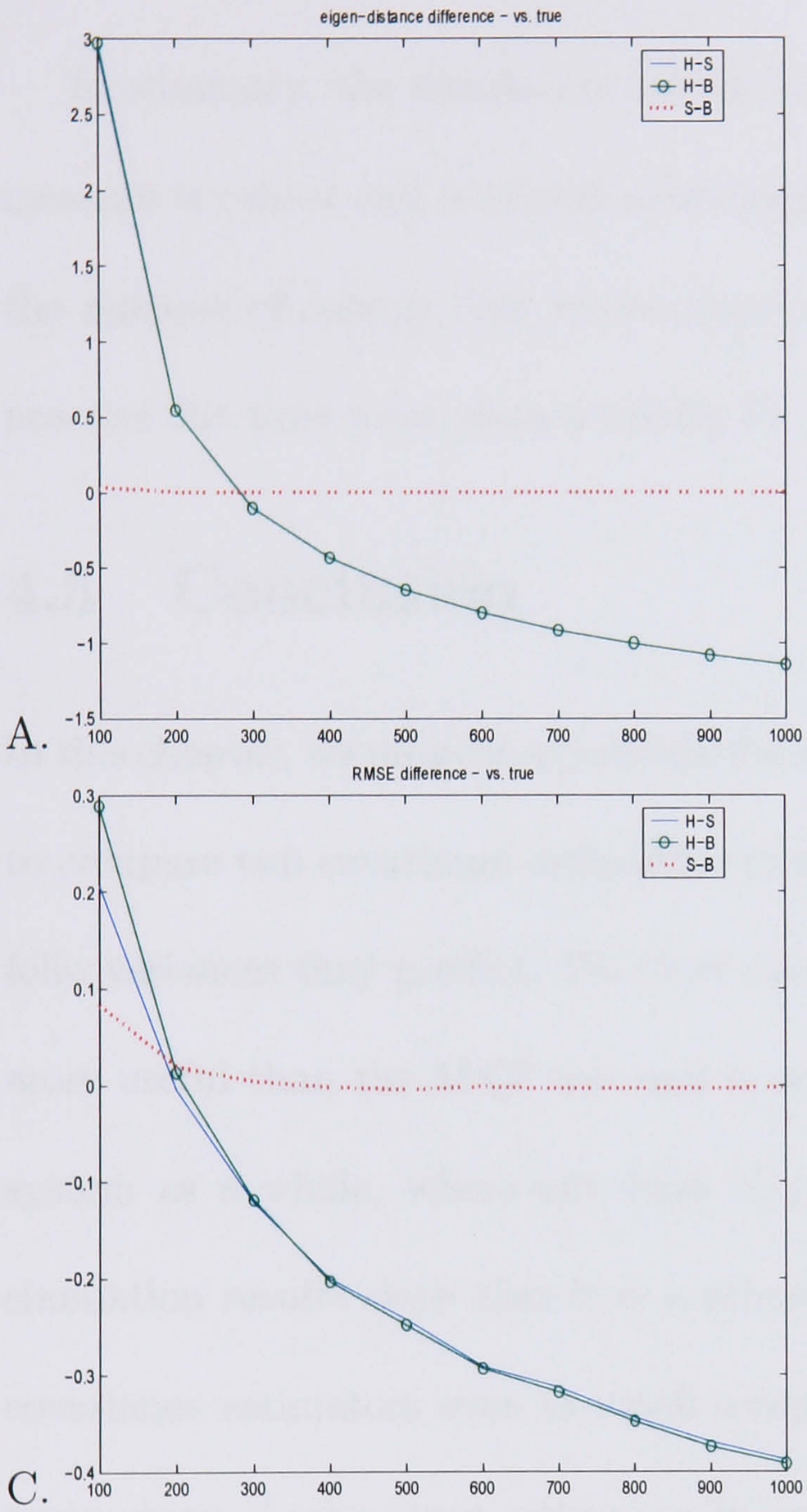


Figure 4.2: **Differences of eigen-distance and RMSE with increasing estimation windows** This figure shows the differences of the eigen-distance and RMSE results of three covariance estimators when the estimation windows increase from 100 weeks to 1,000 weeks. Figure A and B (C and D) plot the eigen-distance (RMSE) results when compared with the *true* covariance matrix and with the *ex-post* covariance matrix respectively. $(H - S)$ represents the difference between the sample historical covariance matrix and the single-index covariance estimator, $(H - B)$ represents the difference between the sample historical covariance matrix and the single-index model with Blume adjusted betas covariance estimator, and $(S - B)$ represents the difference between the single-index estimator and the single-index model with Blume adjusted betas estimator. The horizontal axis is the length of the estimation window in weeks and the vertical axis is the differences of the RMSE or eigen-distance results.

C.

In summary, the simulation results show that our proposed new eigen-distance measure is robust and powerful to distinguish covariance estimators, especially when the amount of sample time series observations is small. This is very helpful as in practice the time series data available for forecasting is usually limited.

4.5 Conclusion

In this chapter, we propose a portfolio distance measure based on eigen-decomposition to compare two covariance estimators in terms of the biggest differences of the portfolio variances they predict. We show that this innovative eigen-distance measure is more useful than the MVP test and is suitable for applying to a risk measurement system as a whole, where any kind of portfolios may need to be measured. Our simulation results show that it is a robust and powerful measure to distinguish two covariance estimators even in small samples, which is very useful for the empirical comparison of covariance estimators in practice.

However, the eigen-distance measure is not very well suited to choose refinements to a particular estimation method. This is the main topic of the next chapter.

Chapter 5

A new measure to assess refinements to risk models

5.1 Introduction

In Chapter 4, we have proposed a portfolio-based eigen-distance measure to compare two covariance matrices based on the biggest differences of the portfolio variances they predict. Our simulation results confirm that it is a useful distance measure. However, it is not suitable to differentiate refinements to a particular covariance estimator.

This chapter introduces a new measure, θ, to determine which one out of two estimated covariance matrices is closer to the observed covariance matrix. It is important in refining a covariance estimator that we know whether the refinement works. More specifically, we can find out if the refined covariance matrix is closer to the observed covariance matrix than is the original covariance matrix. θ is designed to measure

how much two estimated covariance matrices differ and whether one is a clear improvement of the other.

We are particularly interested in a modest refinement, which results in a similar covariance estimator to the original one. We want to find out if the refinement is able to lead to a big improvement on the estimation of a covariance matrix in terms of predicting the extreme portfolio variances.

Suppose $\mathbf{A}$ is the original estimated covariance matrix and $\mathbf{B}$ is similar to $\mathbf{A}$ but incorporates a modest refinement ($\mathbf{A}$ and $\mathbf{B}$ are therefore relatively close together). To differentiate $\mathbf{A}$ and $\mathbf{B}$ from the observed covariance matrix $\mathbf{C}$, we focus on the performance of the x and y portfolios that are predicted to have the most different variances under $\mathbf{A}$ and $\mathbf{B}$, which is the essential difference of $\mathbf{A}$ and $\mathbf{B}$.[1]

We prove that if $\mathbf{A}$ ($\mathbf{B}$) captures the *true* covariance structure, then $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ is expected to be close to 0 (1). However, if neither $\mathbf{A}$ nor $\mathbf{B}$ is close to the true covariance matrix, then $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ is less certain and may be biased although still appear useful under plausible simulation assumptions.[2]

---

[1] x is predicted to be riskier under $\mathbf{A}$ than $\mathbf{B}$, and y is predicted to be safer under $\mathbf{A}$ than $\mathbf{B}$ (similar definitions of x and y as in Chapter 4).

[2] The situation is similar to the specification problem in the OLS estimation. Suppose the *true* model for a stock return is

$$r = a + b_1 f_1 + b_2 f_2 + b_3 f_3 + u \tag{5.1.1}$$

and we have two alternative factor models where

$$r_A = a + b_1 f_1 + u \tag{5.1.2}$$

$$r_B = a + b_1 f_1 + b_2 f_2 + u \tag{5.1.3}$$

We know that because of the specification problem the OLS estimation of b_2 is biased unless f_1 and f_2 are uncorrelated (Theil (1971), Giliberto (1985)).

The rest of this chapter is organized as the following. Section 5.2 defines the new θ measure and derives its analytical properties. Section 5.3 specifies the simulations. Section 5.4 reports the simulation results of θ's sampling properties. Section 5.5 concludes.

5.2 A new θ measure

Assumption 5.2.1. *We assume stock returns are multivariate normal with zero means.*

This assumption makes sure that the returns of any portfolio are normally distributed with zero means. The variances of these portfolios therefore have a chi-square distribution.

Definition 5.2.1. *We define*

$$\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B}) = \frac{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Cx}}{\mathbf{y'Cy}}\right)}{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)} \tag{5.2.1}$$

where $\mathbf{A}$ *and* $\mathbf{B}$ *are two forecasts of the true covariance matrix that are relatively close to each other, and* $\mathbf{C}$ *is an observed ex-post sample from the true covariance matrix.* x *and* y *are two extreme portfolios defined as:*

$$x = Arg\max_{x} \frac{\mathbf{x'Ax}}{\mathbf{x'Bx}} \tag{5.2.2}$$

$$y = Arg\min_{y} \frac{\mathbf{y'Ay}}{\mathbf{y'By}} \tag{5.2.3}$$

so that x *has* more risk *under* $\mathbf{A}$ *than* $\mathbf{B}$, *and* y *has* more risk *under* $\mathbf{B}$ *than* $\mathbf{A}$.

From Equations (5.2.2) and (5.2.3), we know that

$$\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{y}'\mathbf{A}\mathbf{y}} \geq \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{y}'\mathbf{B}\mathbf{y}} \tag{5.2.4}$$

We also assume that the ratio of the two extreme portfolios under the ex-post matrix $\mathbf{C}$ lies between those of the two forecasted covariance matrices, i.e.,

$$\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{y}'\mathbf{A}\mathbf{y}} \geq \frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{\mathbf{y}'\mathbf{C}\mathbf{y}} \geq \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{y}'\mathbf{B}\mathbf{y}} \tag{5.2.5}$$

Or,

$$log\left(\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{y}'\mathbf{A}\mathbf{y}}\right) \geq log\left(\frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{\mathbf{y}'\mathbf{C}\mathbf{y}}\right) \geq log\left(\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{y}'\mathbf{B}\mathbf{y}}\right) \tag{5.2.6}$$

since all portfolio variances are positive. We can see from Equation (5.2.6) that if $\mathbf{A}$ is close to $\mathbf{C}$, $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ will be close to zero, and if $\mathbf{B}$ is close to $\mathbf{C}$, $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ will be close to one. This behavior receives at least some degree of confirmation from the following theorem.

Theorem 5.2.1. *$\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ measures which one of $\mathbf{A}$ and $\mathbf{B}$ is preferred when we observe $\mathbf{C}$. It has the following properties:*

1. *If returns are zero means and multivariate normally drawn from $\mathbf{A}$ (so $\mathbf{C}$ is a random sample of $\mathbf{A}$), then $E[\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})] = 0$; and*

2. *If returns are zero means and multivariate normally drawn from $\mathbf{B}$ (so $\mathbf{C}$ is a random sample of $\mathbf{B}$), then $E[\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})] = 1$.*

Proof of Theorem 5.2.1.

Equation (5.2.1) defines $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ to measure the closeness from $\mathbf{A}$ to $\mathbf{C}$. If we define a new measure $\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ to measure the closeness from $\mathbf{B}$ to $\mathbf{C}$, i.e.,

$$\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B}) = \frac{log\left(\frac{\mathbf{x'Cx}}{\mathbf{y'Cy}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)}{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)} \tag{5.2.7}$$

then,

$$\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B}) + \phi^{\mathbf{C}}(\mathbf{A},\mathbf{B}) = \frac{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Cx}}{\mathbf{y'Cy}}\right)}{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)} + \frac{log\left(\frac{\mathbf{x'Cx}}{\mathbf{y'Cy}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)}{log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)} = 1 \tag{5.2.8}$$

Therefore when $\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ is equal to 0, $\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ must be equal to 1. Equation (5.2.7) can also be written as the following:

$$\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B}) = \frac{log\left(\frac{\mathbf{x'Cx}}{\mathbf{x'Bx}}\right) - log\left(\frac{\mathbf{y'Cy}}{\mathbf{y'By}}\right)}{d(\mathbf{A},\mathbf{B})} \tag{5.2.9}$$

where $d(\mathbf{A},\mathbf{B})$ is the portfolio-based eigen-distance of $\mathbf{A}$ and $\mathbf{B}$ as we have defined in Chapter 4, i.e., $d(\mathbf{A},\mathbf{B}) = \log\left(\frac{\max_x \frac{\mathbf{x'Ax}}{\mathbf{x'Bx}}}{\min_{\mathbf{y}} \frac{\mathbf{y'Ay}}{\mathbf{y'By}}}\right) = log\left(\frac{\mathbf{x'Ax}}{\mathbf{y'Ay}}\right) - log\left(\frac{\mathbf{x'Bx}}{\mathbf{y'By}}\right)$.

If we assume $\mathbf{B}$ captures the true covariance structure and $\mathbf{C}$ is a sample of $\mathbf{B}$, then both $\mathbf{A}$ and $\mathbf{B}$ are fixed (so are the x and y portfolios), $d(\mathbf{A},\mathbf{B})$ is a constant and only $\mathbf{C}$ is a variable. $\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B})$ is affected by the variables in the numerator in Equation (5.2.9): the difference of the two logarithm ratios of the portfolio variances, $log\left(\frac{\mathbf{x'Cx}}{\mathbf{x'Bx}}\right)$ and $log\left(\frac{\mathbf{y'Cy}}{\mathbf{y'By}}\right)$. If we assume a multivariate normal model with zero means, for any fixed vector x s.t. $x'\mathbf{B}x > 0$

$$n \times \frac{x'\mathbf{C}x}{x'\mathbf{B}x} \sim \lambda_n^2 \tag{5.2.10}$$

and similarly, for any fixed vector y s.t. $y'\mathbf{B}y > 0$

$$n \times \frac{y'\mathbf{C}y}{y'\mathbf{B}y} \sim \chi^2_n \tag{5.2.11}$$

where n is the degree of freedom for the chi-squared distribution. n also equals the number of the weekly time-series observations of the portfolio returns (or the stock returns).[3]

We can derive that the expected return and standard deviation of the log chi-square distribution are[4]

$$E\left(\log \chi^2_n\right) \simeq \log(n) \tag{5.2.12}$$

$$SD\left(\log \chi^2_n\right) \simeq \frac{1}{\sqrt{(n-1)/2}} \tag{5.2.13}$$

Since $log\left(\frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}}\right)$ and $log\left(\frac{\mathbf{y}'\mathbf{C}\mathbf{y}}{\mathbf{y}'\mathbf{B}\mathbf{y}}\right)$ are from the same distribution determined by n, we have

$$E\left[\phi^{\mathbf{C}}(\mathbf{A},\mathbf{B})\right] = 0 \tag{5.2.14}$$

---

[3]This can be easily proved. Since returns of stocks are assumed to be multivariate normal with zero means, the return of any portfolio is also normally distributed with zero mean. The variance of the portfolio is therefore distributed as a chi-square distribution, and the degree of freedom equals the number of portfolio returns (Johnson and Kotz (1972)). Since $x'\mathbf{C}x$ $(y'\mathbf{C}y)$ is the variance of portfolio x (y), it follows a chi-squared distribution.

[4]The moment generating function of a log gamma distribution (X) is

$$E\left(e^{t\log X}\right) = E\left(X^t\right) = \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}$$

Except for very small sample sizes, we can use the digamma and trigamma functions to calculate the mean and the standard deviation (SD) of the log gamma distribution, which approximately equal

$$E\left(\log X\right) \simeq \log\alpha$$

$$SD\left(\log X\right) \simeq \frac{1}{\sqrt{\alpha - 1/2}}$$

This proves Result 1. According to Equation (5.2.8), we also have

$$E\left[\theta^{\mathbf{C}}(\mathbf{A},\mathbf{B})\right] = 1 \tag{5.2.15}$$

This proves Result 2.

This completes the proof of Theorem 5.2.1..

5.3 Simulations

Theorem 5.2.1. does not provide any indication of the power of θ to differentiate between two covariance matrices. Furthermore, if neither $\mathbf{A}$ nor $\mathbf{B}$ is the true covariance structure, then the θ measure may be biased. We will use simulations to illustrate these two issues. This section describes the specifications for our simulations.

5.3.1 Specifying two covariance estimators

We denote the forecasted covariance matrix obtained using the single-index model as $\mathbf{A}$ and the forecasted covariance matrix obtained using the single-index model with Blume adjusted betas as $\mathbf{B}$. As a result, $\mathbf{B}$ is similar to $\mathbf{A}$ but incorporates some refinement.

We consider two scenarios. The first scenario is to justify the theorem when one of $\mathbf{A}$ and $\mathbf{B}$ is close to the true covariance matrix. The second scenario explores the usefulness of θ when both $\mathbf{A}$ and $\mathbf{B}$ are a long way from the true covariance matrix. We study if θ can better distinguish these two covariance estimators than both the

eigen-distance and the RMSE do.

5.3.2 One estimator captures the true covariance structure

In this case, we assume that $\mathbf{B}$ captures the true covariance structure and $\mathbf{C}$ is a sample from $\mathbf{B}$. We estimate $\mathbf{A}$ and $\mathbf{B}$ using the Wednesday to Wednesday weekly stock prices of the 78 NYSE US 100 Index component companies from 01.1988 to 12.1993. The S&P 500 index for the corresponding period is used as the market index. The empirical data is divided into two equal length sub-sample periods: 01.1988-12.1990 and 01.1991-12.1993.

$\mathbf{C}$ is constructed using the artificial returns generated from $\mathbf{B}$ according to the decomposition method introduced in Chapter 4. More specifically, we assume the returns follow a N-variate factor model, where the ith factor has zero mean and variance equal to the ith eigenvalue of the covariance matrix $\mathbf{B}$. The factor loading of ith factor on individual stock j equals the jth element of the ith eigenvector matrix of the covariance matrix $\mathbf{B}$. We expect θ to be statistically different from 0 but not statistically different from 1.[5]

[5]The opposite is that if $\mathbf{A}$ captures the true covariance matrix, then $\theta^{\mathbf{C}}(\mathbf{A}, \mathbf{B})$ will be statistically close to 0 but different from 1.

5.3.3 Neither estimator is close to the true covariance structure

In the second simulation scenario, we assume that neither **A** nor **B** is close to the *true* covariance matrix. We assume the true covariance matrix contains a richer structure as defined in Chapter 4 (Section 4.3.1). In this case, all **A**, **B** and **C** are constructed using the artificial returns simulated from the *true* covariance matrix: **A** and **B** are constructed according to the single-index model and the single-index model with Blume adjusted betas, and **C** is a sample of the true covariance matrix. We expect θ to be less certain than it is in the first simulation scenario, and may be biased and unable to provide the correct signal. We show that θ may still appear to be useful under plausible simulation assumptions.

In both simulations, we assume the true covariance structure is constant. We vary the length of the time-series observations to check the robustness of the results. We consider 200, 600 and 1000 weeks. We also increase the simulations to 1,000 times from 100 times used in Chapter 4 to obtain more significant results.

5.4 Results

Table 5.1 reports the average of 1000 simulations of the comparison results of **A** and **B** against a sample *true* covariance matrix **C** for 200, 600 and 1000 weeks of time series observations respectively. We report four types of results: the RMSE differences, the

eigen-distance differences, θ and the average beta adjustment (b) in the single-index model with Blume-adjusted betas.[6]

Panel A of Table 5.1 reports the results when one of the two covariance matrices, **B**, captures the true covariance matrix. The 10th (11th) column reports the statistical significance of θ being different from 0 (1), or if **C** is significantly different from **A** (**B**). We find that θ is statistically significantly different from 0 but not statistically significantly different from 1. This shows that **C** covariance matrix is significantly closer to **B** than to **A**. For example when T=200, θ equals 0.9995 and *t-statistic against 0* equals 122.11 and *t-statistic against 1* equals -0.06. We also find that the significance of the t-statistics increases since the standard errors reduce with increasing time-series observations.

In addition, we find that compared to the differences of RMSE and eigen-distance, θ results are a lot more significant. For example, when T=200, the t-statistics for the differences of the RMSE and eigen-distance equal 30.55 and 21.10 respectively, while the *t-statistic against 0* of θ equals 122.11.

---

[6]Using the notations in Chapter 3, we have $\beta_{2,i} = a + b\beta_{1,i} + c_i$ and $\beta_{Blume,i} = a + b\beta_{2,i}$. b measures the amount of unadjusted betas in the adjusted betas. The higher the b, the smaller the adjustment is, as the difference between the adjusted and unadjusted betas is smaller.

Table 5.1: **RMSE, eigen-distance, $\theta^{C}_{A,B}$ and β adjustment** This table reports the comparison results of the covariance matrices obtained according to the *single-index model* (**A**) and *single-index model with Blume-adjusted betas* (**B**) against a sample of the true covariance matrix (**C**). The reported results are based on 1,000 simulations when the time-series observation (T) equals 200, 600 and 1,000 weeks. The *RMSE difference* (A-B) and *eigen-distance difference* (A-B) measure the average pair-wise differences of the RMSE ($RMSE(A) - RMSE(B)$) and the eigen-distance of **C** with **A** and **B** ($d(\mathbf{A},\mathbf{C}) - d(\mathbf{B},\mathbf{C})$) respectively. The reported β adjustment (b) equals the average weight of unadjusted betas in the Blume adjusted betas. Panel A reports the results when we assume that **B** captures the actual covariance matrix and **C** is a sample of **B**. Panel B reports the results when neither **A** nor **B** is close to the true covariance matrix. The $*$ indicates that the numbers are statistically significant at a 95% confidence level.

	RMSE(A)-RMSE(B)			$d(\mathbf{A},\mathbf{C}) - d(\mathbf{B},\mathbf{C})$			$\theta^{\mathbf{C}}_{\mathbf{A.B}}$	se	t-stats	t-stats	β adjustment
	mean	se	t-stats	mean	se	t-stats			against 0	against 1	(b)
Panel A: Scenario 1. **B** close to **C**											
T=200	0.1573	0.0051	(30.55 *)	0.0109	0.0005	(21.10 *)	0.9995	0.0082	(122.11 *)	(-0.06)	0.97
T=600	0.2409	0.0045	(53.22 *)	0.0326	0.0009	(37.74 *)	0.9941	0.0046	(215.52 *)	(-1.27)	0.97
T=1,000	0.2991	0.0039	(76.86 *)	0.0603	0.0010	(59.45 *)	1.0061	0.0036	(280.20 *)	(1.70)	0.97
Panel B: Scenario 2. neither **A** nor **B** close to **C**											
T=200	0.0122	0.0012	(10.00 *)	-0.0011	0.0005	(-2.08 *)	0.8513	0.0725	(11.74 *)	(-2.05 *)	0.84
T=600	-0.0040	0.0005	(-7.76 *)	-0.0005	0.0002	(-3.40 *)	0.3075	0.1400	(2.20 *)	(-4.95 *)	0.94
T=1,000	-0.0039	0.0003	(-12.44 *)	-0.0002	0.0001	(-2.03 *)	0.2030	0.1098	(1.85)	(-7.26 *)	0.96

We also find that the average Blume adjustment on betas is fairly small (as b of 0.97 is quite big), indicating a small difference in the adjusted and unadjusted betas, and close covariance estimators for the single-index model and the single-index model with the Blume-adjusted betas. Even so, our θ measure is still able to clearly differentiate them, including in a small sample of time-series observations. For example when T=200, the average θ equals 0.9995 with standard deviation equals 0.259 (=0.0082 $\times$ $\sqrt{1,000}$). We know that the sample size of the time-series observations for forecasting purposes is usually small. The finding that θ performs very well in small samples is particularly useful.

Panel B reports the results when neither **A** nor **B** is close to a richer true covariance structure. Overall, we find that comparison results (θ, eigen distance and RMSE) regarding the differences of **A** and **B** are less significant compared to those under Scenario 1 where **B** is assumed to capture the true covariance structure. For example, the t-statistics for θ is 11.74 instead of 122.11 for T=200.

When T=200, θ results still indicate that **C** is closer to **B** than to **A**, but they are less certain than those under Scenario 1. For example, the average θ equals 0.8513 and is statistically significantly different from both 0 (*t-statistics against 0* equals 11.74) and 1 (*t-statistics against 1* equals 2.05).

When the size of sample time-series observations increases, we find that the average θ reduces and the results of its t-statistics against 0 and 1 reverse, which indicates

that **C** is getting closer to **A** than to **B**. More specifically, θ reduces from 0.8513 to 0.3075 and 0.2030 when T increases from 200 to 600 and 1000. The corresponding t-statistics against 0 decrease from 11.74 to 2.20 and 1.85, and the t-statistics against 1 increase from 2.05 to 4.95 and 7.26. These results seem to indicate that the Blume adjustment on betas does not work when the neither single-index covariance estimators are close to the true covariance matrix, especially when the sample size is big.

In terms of the Blume adjustment on betas, we find that the adjustment is smaller when the sample size increases. More specifically, the average adjustments on betas for the three samples equal 0.84, 0.94 and 0.96 respectively.[7] It shows that when the amount of time-series observations increases, there is less difference between **A** and **B**. Given that we also assume both **A** and **B** are a long way from the observed covariance matrix **C**, it is difficult to choose between these two covariance matrices. The θ results may be biased and give the wrong signal.

Table 5.2 reports the pair-wise eigen-distances of **A**, **B** and **C**, and $d(\mathbf{A}, \mathbf{B})$ as a fraction of $d(\mathbf{A}, \mathbf{C})$ and $d(\mathbf{B}, \mathbf{C})$ for both Scenario 1 and 2. We can see in Scenario 2 (Panel B) that when the time-series observations increase, $d(\mathbf{A}, \mathbf{B})$ (the eigen-distance of the two estimated covariance matrices) reduces at a faster rate than both $d(\mathbf{A}. \mathbf{C})$ and $d(\mathbf{B}, \mathbf{C})$ (the eigen-distance of the estimated covariance matrices to the observed

[7]This is expected as the standard error of betas decreases as time-series observations increase (Vasicek (1973)).

Table 5.2: **Relative distances to the ex-post covariance matrix** This table reports the average pair-wise eigen-distances of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ over 1000 simulations, for time-series observations equal 200, 600 and 1000 weeks respectively. The number in bracket represents the fraction of $d(\mathbf{A},\mathbf{B})$ to $d(\mathbf{A},\mathbf{C})$ and $d(\mathbf{B},\mathbf{C})$ respectively.

	$d(\mathbf{A},\mathbf{B})$	$d(\mathbf{A},\mathbf{C})\left(\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{A},\mathbf{C})}\right)$	$d(\mathbf{B},\mathbf{C})\left(\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{B},\mathbf{C})}\right)$
Panel A: Scenario 1. $\mathbf{B}$ close to $\mathbf{C}$			
T=200	0.55	2.8335 (0.19)	2.8227 (0.19)
T=600	0.55	1.4884 (0.37)	1.4558 (0.38)
T=1,000	0.55	1.1672 (0.47)	1.1069 (0.49)
Panel B: Scenario 2. neither $\mathbf{A}$ nor $\mathbf{B}$ close to $\mathbf{C}$			
T=200	0.51	3.6832 (0.14)	3.6843 (0.14)
T=600	0.20	2.6695 (0.08)	2.6700 (0.08)
T=1,000	0.14	2.4908 (0.06)	2.4910 (0.06)

covariance matrix). As a result, the relative distances of both $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{A},\mathbf{C})}$ and $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{B},\mathbf{C})}$ reduce from 0.14 to 0.08 and 0.06. As we find that θ does not work very well for T=600 and 1000 in Scenario 2, we suspect that θ only works for a certain range of the relative distance of $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{A},\mathbf{C})}$ and $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{B},\mathbf{C})}$. According to Table 5.2, it appears that $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{A},\mathbf{C})}$ or $\frac{d(\mathbf{A},\mathbf{B})}{d(\mathbf{B},\mathbf{C})}$ needs to be at leat more than 0.15, in order for θ to have significant and robust results.

To summarize, we find that θ is more powerful than the RMSE and eigen-distance measures to differentiate between two covariance matrices when they are relatively close together. However, if these matrices are a long way from the true covariance structure, the θ measure (not surprisingly) may give unhelpful signals.

5.5 Conclusion

The general problem we wish to address in this chapter is to differentiate between two similar alternative ways of estimating the covariance matrix when we observe a sample from the actual covariance matrix. We introduce a θ measure which is more powerful than both the RMSE and eigen-distance measures to differentiate these close covariance matrices. This θ measure is very helpful in refining covariance estimators.

Our simulation results show that the relative power of θ depends on the sampling size as well as the relative closeness of the two matrices and their distances to the actual covariance matrix. For two reasonably close covariance matrices, θ is useful even in small samples. When neither of the two covariance matrices is close to the true structure, the behavior of θ becomes less certain (and it may be biased), but still appears useful under plausible simulation assumptions.

Chapter 6

Improving estimation of the covariance matrix of stock returns

6.1 Introduction

We know that a large-scale sample covariance matrix based on historical data often contains large estimation errors as the amount of time series observations is often not big enough compared to the number of parameters required to estimate for the covariance matrix. When we impose more structure on the covariance matrix to reduce the amount of parameter estimations (and hence the estimation errors), we also introduce specification errors as the structural assumptions are usually only approximately valid.

As we have seen in Chapter 2, in recent years, several studies have concentrated on obtaining a trade-off between estimation errors and specification errors through linear combinations of the sample covariance matrix and the single-index model covariance

estimator (or a constant correlation estimator).[1] These studies find that taking a weighted average of covariance estimators provides a better estimate of the covariance matrix than any single estimator. The question is what kind of weighting schemes we should apply to the estimators. We first briefly revisit some of the key papers.

Ledoit and Wolf (2003a) suggest an empirical Bayesian shrinkage approach that involves solving a quadratic loss function based on the Frobenius norm to find the optimal weights on the sample covariance matrix and the single-index estimator. They find that the optimal weight on the single index estimator equals to $\hat{\alpha}^* = f(\hat{\kappa})$ where $\hat{\kappa} = \frac{\hat{\pi}-\hat{\rho}}{\hat{\gamma}}$ and show that it is a consistent estimator.[2] However, this large sample property is found not to be very useful as a large amount of time series observations in practice is not realistic. In addition, although the authors prove that $\hat{\pi}$, $\hat{\rho}$ and $\hat{\gamma}$ are consistent, their ratio $\hat{\kappa}$ could still be biased.

Jagannthan and Ma (2003) examine whether a simple average of the single-index estimator and the sample covariance matrix would do equally well since the Ledoit and Wolf estimator is a particular weighted average of the two estimators. They find that the simple average works as well as the Ledoit and Wolf's shrinkage estimator.

Recently, Disatnik and Benninga (2005) use empirical data to compare several shrinkage estimators and portfolios of estimators (simple average of a number of different covariance estimators) under the ex-post global minimum variance portfolio

---

[1] See for example, Daniels and Kass (2001), Ledoit and Wolf (2003a, 2003b, 2004), Schafer and Strimmer (2005), Jagannathan and Ma (2003), Bengtsson and Holst (2002).

[2] See Chapter 3 for a brief review. For details, refer to Ledoit and Wolf (2003a).

criterion and find it is impossible to claim any one method to be better than the other. They conclude that there is no additional benefit for using more sophisticated shrinkage methods than a simpler portfolio of estimators.

In this chapter, we are interested in improving the best available covariance estimators. We explore some alternative covariance estimators based on directly shrinking the eigenvalues of the sample covariance matrix. We use simulations to compare their performances with the Ledoit and Wolf (2003a) estimator and the Jagannathan and Ma (2003) estimator using the RMSE and the portfolio-based eigen-distance criteria. Contrary to the empirical studies, as the *true* covariance matrix structure is exactly known and we can control the amount of time series observations, we can compare more precisely the difference of various covariance estimators and the effect of different lengths of the sample period on the performance ranking of alternative estimators.

We find that our shrinkage estimators consistently out-perform the Ledoit and Wolf (2003a) estimator. They also perform better than the Jagannathan and Ma estimator by reasonable amounts with only one exception when the amount of time-series data is relatively small, where the Jagannathan and Ma estimator performs better than our shrinkage methods. However, even in this case, our shrinkage methods are not a lot worse than the Jagannathan and Ma estimator. We also find that the choice of comparison criteria affects the relative ranking of the covariance estimators.

Our shrinkage methods rank more favorably under the RMSE criterion than under the eigen-distance measure, especially for small amount of time-series data.

The remaining of the chapter is organized as the following. Section 6.2 describes our alternative shrinkage estimators. Section 6.3 discusses the simulation results. Section 6.4 concludes.

6.2 Improving a sample covariance matrix

6.2.1 Eigenvalue shrinkage methods

It is known that the distribution of a sample covariance matrix S of the covariance matrix Σ of a multivariate normal population can be described by a Wishart distribution as $S \sim W(\Sigma, n)$, where $S = \Sigma/n$ and n is the degree of freedom. The eigenvalues of the sample covariance matrix S tend to be much more dispersed than the eigenvalues of the population covariance matrix Σ, and the excess dispersion equals the error of the sample covariance matrix. In other words, the bigger eigenvalues are more likely to contain positive measurement errors and the smaller eigenvalues to contain negative measurement errors.

An intuitively appealing approach to improve the sample covariance matrix is to shrink the sample eigenvalues towards some central value (Muirhead (1987)). Daniels and Kass (2001) provide an extensive review of the empirical Bayesian shrinkage estimators (where the empirical data determine the amount of shrinkage) proposed

in recent years. They classify these methods into two general approaches. The first approach involves shrinking the eigenvalues of the sample covariance matrix and the second approach involves shrinking the unstructured estimator (the sample covariance matrix) towards a structured estimator (also called a prior or a shrinkage target). The Ledoit and Wolf (2003a)'s shrinkage method falls into the second approach, which could also be regarded as a more complicated eigenvalue shrinkage method.

6.2.2 Four simple eigenvalue shrinkage methods

Several studies have used orthogonally invariant estimators of the form $\tilde{\mathbf{\Sigma}} = \mathbf{T}\mathbf{\Lambda}^*(\hat{\lambda})\mathbf{T}^\mathbf{T}$, where T is the matrix of normalized eigenvectors, $\hat{\lambda}$ is the vector of sample eigenvalues, and $\Lambda^*(\hat{\lambda}) = diag(\lambda_1^*(\hat{\lambda}), \ldots, \lambda_p^*(\hat{\lambda}))$, where each λ_j^* is a real-valued non-negative function.[3]

In this section, we describe four relatively simple eigenvalue shrinkage methods. Our proposed eigenvalue shrinkage methods impose minimum structure assumptions during the adjustment process, in order to reduce the potential specification errors introduced while trying to improve the sample eigenvalues. The four methods are described in the order of increasing technical complexity.

The first method, *Eigenvalue adjustment* (V for short), assumes orthogonality of eigenvectors and uses the eigenvalues of one sample covariance matrix to adjust

[3]See for example, Stein (1975), Haff (1980, 1991) Dey and Srinivasan (1985), and Yang and Berger (1994).

the eigenvalues of a sample covariance matrix obtained from a prior sample period. This method is inspired by the Blume's (1971) linear adjustment on the estimates of betas of the single-index model, as the estimated sample eigenvalues behave like the estimated sample betas, in that the higher eigenvalues have positive measurement errors and the lower eigenvalues have negative measurement errors. This method has appeared earlier in the thesis to construct the *true* covariance matrix for our simulation in Chapter 4. More specifically, we first obtain two covariance matrices for two successive sample periods $\mathbf{C_1}$ and $\mathbf{C_2}$. We calculate the eigenvalue matrix $\mathbf{\Lambda_1}$ and eigenvector matrix $\mathbf{T}_1$ of $\mathbf{C_1}$ using eigen-decomposition $\mathbf{C_1 = T_1 \Lambda_1 T_1'}$. We then adjust the eigenvalues using $\mathbf{T_1' C_2 T_1 = \Lambda}$ and form a new diagonal eigenvalue matrix $\mathbf{\Lambda^*}$ by keeping only the diagonal values of $\mathbf{\Lambda}$. The adjusted covariance matrix is then constructed as[4]

$$\mathbf{C_V = T_1 \Lambda^* T_1'} \tag{6.2.1}$$

Our second method, *Smoothed-eigenvalues* (*SV* for short), smoothes the adjusted eigenvalues $\mathbf{\Lambda^*}$ produced by the *Eigenvalue adjustment* estimator using a cubic spline smoothing method, and uses the smoothed eigenvalue matrix $\mathbf{\Lambda^*_{smoothed}}$ to construct the covariance estimation.

$$\mathbf{C_{SV} = T_1 \Lambda^*_{smoothed} T_1'} \tag{6.2.2}$$

We find that the first few eigenvalues are always a lot larger than the remaining

[4]For details refer to Chapter 4.

eigenvalues.[5] We learn from random matrix theory that these eigenvalues are likely to contain real information regarding the structure of the covariance matrix. We therefore do not smooth them. As a smoothing rule, we spare the first five adjusted eigenvalues and smooth the remaining adjusted eigenvalues. We also specify the smoothing parameter to be 0.01 so that we have a relatively smoothed plot of the high order adjusted eigenvalues (not reported).[6]

The third method, *Smoothed-log-eigenvalues* (*SVlog* for short), is similar to the SV method except that we smooth the high order log-eigenvalues instead of the absolute eigenvalues themselves. The intension is to reduce the contribution from the noisy part of the covariance matrix.

The fourth method, *Eigenvalue adjustment with Blume-adjusted-eigenvector* (*TV* for short), adjusts the first eigenvector ($\mathbf{t}_{1,1}$) in addition to adjusting the eigenvalues. The adjusted covariance matrix is obtained as

$$\mathbf{C_{TV}} = \mathbf{T_{TV}} \mathbf{\Lambda}^* \mathbf{T}'_{\mathbf{TV}} \tag{6.2.3}$$

where $\mathbf{T}_{TV} = [\mathbf{t}_{1,Blume}, \mathbf{t}_{1,i}, \ldots \mathbf{t}_{1,N}]$ and $\mathbf{t}_{1,i}$ is the ith ($i \in [2, N]$) eigenvector of the eigenvector matrix $\mathbf{T}_1$. $\mathbf{t}_{1,Blume}$ is the Blume-adjusted first eigenvector. As the first eigenvector is similar to the loadings on the market factor (or betas), and it is well

[5] An example of this is illustrated by the results reported in Table 4.1 in Chapter 4.

[6] The cubic spline smoothing requires balancing two opposing criteria (1) the spline must come reasonably close to the data; and (2) the spline must also have low curvature. We use the Matlab cubic spline smoothing where the smoothing spline minimizes $p \cdot \Sigma_{i=6}^{n} [x_i - y_i]^2 + (1-p) \cdot \int |x''|^2 dx$, x'' being the second derivatives of x. p is the smoothing parameter and determines the relative weight on the contradictory requirements. p is between 0 and 1: $p = 0$ produces a least square straight line fit to the data, while $p = 1$ produces a cubic spline interpolant.

known that the Blume adjustment helps reduce the biases of betas in the single-index model, we apply a similar Blume adjustment on the first eigenvector.

The reason for adjusting only the first eigenvector is that $\mathbf{t}_{1,1}$ is the only eigenvector whose mean is clearly different from zero. Therefore, adjusting the elements of the first eigenvector towards their mean will have a different effect from adjusting the corresponding eigenvalue. On the other hand, the higher order eigenvectors have means more close to zeros, in which case adjusting their elements towards zeros has the same effect as adjusting their corresponding eigenvalues. Since we have already adjusted the eigenvalues, we do not need to adjust the elements of high order eigenvectors.

To carry out the Blume adjustment, we need to find the principal eigenvectors for two successive sample periods. If we define $\mathbf{R}_{m,1}$ and $\mathbf{b}_1$ as a scalar multiple of the market return and the stock betas,[7] we have $\mathbf{R}_{m,1} = \mathbf{R}_1\mathbf{t}_{1,1}$ and $(\mathbf{V}_1 - \lambda_1\mathbf{I})\mathbf{t}_{1,1} = 0$, therefore $\mathbf{b}_1 = \frac{cov(\mathbf{R}_1,\mathbf{R}_{m,1})}{var(\mathbf{R}_{m,1})} = \frac{\mathbf{R}_1'\mathbf{R}_1\mathbf{t}_{1,1}}{\mathbf{t}_{1,1}'\mathbf{R}_1'\mathbf{R}_1\mathbf{t}_{1,1}} = \frac{\mathbf{V}_1\mathbf{t}_{1,1}}{\mathbf{t}_{1,1}'\mathbf{V}_1\mathbf{t}_{1,1}} = \mathbf{t}_{1,1}$, where $\mathbf{R}_1$ and $\mathbf{V}_1$ are the return matrix and the covariance matrix for the first sample period respectively. If we have the covariance matrix for the second sample period, $\mathbf{V}_2$, betas for the second sample period can be obtained using $\mathbf{b}_2 = \frac{\mathbf{V}_2\mathbf{t}_{1,1}}{\mathbf{t}_{1,1}'\mathbf{V}_2\mathbf{t}_{1,1}}$. We then apply the standard Blume adjustment method on the betas, i.e.,

$$\mathbf{b}_{2,j} = a + b_{\mathbf{t}_{1,1}}\mathbf{b}_{1,j} + e_j \tag{6.2.4}$$

$$\mathbf{t}_{1,Blumej} = a + b_{\mathbf{t}_{1,1}}\mathbf{b}_{2,j} \tag{6.2.5}$$

[7] Because $\sum t_{1,1i}^2 = 1$ but $\sum t_{1,1i} \neq 1$, therefore $\mathbf{R}_1\mathbf{t}_{1,1}$ is not exactly the market return.

where j denotes the jth stock ($j \in [1, N]$), and $b_{\mathbf{t}_{1,1}}$ indicates the amount of unadjusted eigenvector in the adjusted eigenvector.

6.3 Results

Table 6.1 - Table 6.6 report the average of 100 simulations[8] of the RMSE and eigen-distances (in Panel As) and the pair-wise differences of the covariance estimators (in Panel Bs) against the true covariance matrix and an ex-post sample of the true covariance matrix when the length of time-series observations equals 200, 600 and 1,000 weeks.

First of all, we find that our four shrinkage methods consistently out-perform all other estimators in almost all reported cases with one exception under the eigen-distance measure for T=200 (Table 6.2) where they are found to slightly under-perform the Jagannathan and Ma estimator (JM). Even so, they are not a lot worse than the JM estimator. When compared against the true covariance matrix, the difference between SV and JM is only about 14 percent (=(1.6849-1.6431)/(1.9861-1.6868)) of the difference between our shrinkage methods and the next best method (i.e., SV versus the Ledoit and Wolf estimator).

[8]We can obtain more precise estimations of different covariance estimators as well as their differences by increasing the number of simulations. The reason that we do not use more simulations here is that we find the standard errors (hence the standard deviations) of both the RMSE and eigen-distance results for different estimators are very close. Therefore, we can compare estimators through their mean differences. The differences of estimators do not have similar standard errors, as some estimators (especially our four shrinkage estimators) are more highly correlated than others.

We find that although Ledoit and Wolf (2003a, 2004) show their optimal shrinkage intensity estimator is consistent, this asymptotic property turns out not to be very useful. The Ledoit and Wolf estimator only performs better than the Jagannathan and Ma estimator for quite large T, when it is also close to the historical matrix.[9] The Ledoit and Wolf estimator places more weight on the sample covariance matrix when the amount of time series observations increase. As the sample historical matrix performs better with more observations, the Ledoit and Wolf estimator performs better than the Jagannathan and Ma estimator, which always has a constant weight (50%) on the sample covariance matrix and relatively under-weights the sample historical covariance matrix.

The Ledoit and Wolf estimator also consistently under-performs our four shrinkage estimators under both the RMSE and eigen-distance measure. This suggests that rather than using the asymptotic optimal weight proposed by the Ledoit and Wolf (2003a, 2003b, 2004), it would be better to find an alternative simple empirical estimation approach.

Among our four shrinkage estimators, we find that smoothing log-eigenvalues (SVlog) on average does better than smoothing the eigenvalues themselves (SV). However, the overall usefulness of the smoothing techniques reduces with increasing

---

[9]The average shrinkage intensity for the Ledoit and Wolf estimator, or the weight they assign to the single-index covariance estimator equal 0.20, 0.08 and 0.05 for 200, 600, and 1000 weeks and their corresponding standard deviations equal 0.02, 0.006 and 0.003. That is, when T=1,000 weeks, the Ledoit and Wolf estimator places 95% of weight on the sample historical covariance matrix.

time-series observations. More specifically, when T increases, both the SV and SVlog estimators are increasingly worse off than the V estimator, which does not smooth the eigenvalues. For example, the difference between the SV estimator and the V estimator, $(SV - V)$, equals to -0.0019 (with t-statistic of -6.72) when T=200, but increases to 0.0058 (with t-statistic of 27.71) when T=1,000.

In addition, we do not find evidence that the TV estimator consistently outperforms the V estimator. In other words, adjusting the principal eigenvector does not seem to add additional benefits if we are already adjusting the eigenvalues. This shows that eigenvalues have more influence on the stability of the covariance matrix than the eigenvectors, which probably explains why there are established methods of adjusting the eigenvalues rather than eigenvectors. It may also be that our method of adjustment is sub-optimal.[10]

We calculate the pair-wise eigen-distance of two estimated covariance matrices (not reported). If it is relatively small and the difference of the eigen-distances to the true or ex-post covariance matrix reported in Table 6.1 is also not statistically significant, we calculate the corresponding θ results to see if we could find more significant differences of the two estimators that the eigen-distance measure has not identified. We find that there are two pairs of estimators that are relatively close together. The first pair is the SV estimator and the Jagannathan and Ma estimator

[10] We find that average adjustment on the first eigenvector, $b_{\mathbf{t}_{1,1}}$ (as in Equation (6.2.5)), equals to 0.85, 0.93 and 0.95 for T=200, 600 and 1,000 weeks respectively.

when T=200 and compared to the ex-post covariance matrix. We find that θ for (SV-JM) equals to 0.9733 and t-statistics against 0 and 1 equal to 52.66 and -1.45 respectively. It shows that in this case the JM estimator is significantly closer to the ex-post covariance matrix than is the SV estimator. The second pair is the TV estimator and the V estimator for all three reported lengths of the time-series observations. We find that these two estimators are too close together relative to either the true or ex-post covariance matrix and θ is not able to provide any useful indications regarding their relative performance.

We find not surprisingly that the RMSE and eigen-distance results of all estimators as well as the corresponding estimation errors decrease with the increasing amount of time-series observations. This shows that if the world is constant, all estimation methods are more precise with more observations. In addition, we find that the rankings of estimators are different under the RMSE and eigen-distance when the sample time-series observations are small, but they become more consistent when the sample observations increase. This shows that the choice of comparison criteria affects the relative performance rankings of the covariance estimators for a small sample.

To summarize, we find that our four eigenvalue shrinkage estimators consistently out-perform the best available estimators (the Ledoit and Wolf estimator and the Jagannathan and Ma estimator) by a considerable amount under both the RMSE

and eigen-distance measure, except in one case. When T=200 and under the eigen-distance measure, the Jagannathan and Ma estimator only performs slightly better than our simpler shrinkage estimators.

6.4 Conclusion

In this chapter, we have attempted to improve on the best methods of estimating a covariance matrix within the literature and we have succeed to a considerable extent. We explore a few new estimators based on Bayesian shrinkage to the eigenvalues (and in one case to the principal eigenvector as well), and compare their performance with the Ledoit and Wolf estimator and the Jagannathan and Ma estimator using both the RMSE and eigen-distance criteria. We find that our proposed estimators consistently beat the Ledoit and Wolf estimator. They also beat the Jagannathan and Ma estimator by a reasonable amount except in one case, where they are not a lot worse off than the Jagannathan and Ma estimator.

All our simulations until now have been based on the assumption of an unchanging multivariate normal world. In next chapter, we will explore the implications of both fat tails and time variation in stock returns.

Table 6.1: **RMSE and RMSE differences (T=200 weeks)** This table reports the 100 times simulation results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 200 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. The smaller number means the estimator is better. Results in Panel B are reported in the fixed order. The short forms used are: H for the sample historical estimator, S for the single-index covariance estimator, LW for the Ledoit and Wolf estimator, JM for the Jagannathan and Ma estimator, V for the Eigenvalue-adjustment estimator, SV for the Smoothed-eigenvalues estimator, $SVlog$ for the smoothed-log-eigenvalues estimator, TV for the Blume-adjusted-first-eigenvector plus eigenvalue-adjustment estimator. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
S	1.0867	0.0108	S	1.5503	0.0164
H	1.0800	0.0095	H	1.5489	0.0166
LW	0.9542	0.0095	LW	1.4598	0.0181
JM	0.9209	0.0114	V	1.4439	0.0189
V	0.9130	0.0094	SV	1.4427	0.0189
SV	0.9111	0.0094	SVlog	1.4422	0.0189
SVlog	0.9105	0.0094	JM	1.4407	0.0173
TV	0.9036	0.0114	TV	1.4259	0.0179
Panel B: RMSE Differences					
	mean	se t-stats		mean	se t-stats
S - H	0.0067	0.0042 (1.61)	S - H	0.0014	0.0053 (0.26)
LW - H	-0.1258	0.0028 (-45.27*)	LW - H	-0.0891	0.0038 (-23.31*)
LW - S	-0.1325	0.0054 (-24.49*)	LW - S	-0.0905	0.0074 (-12.20*)
JM - H	-0.1591	0.0032 (-49.07*)	JM - H	-0.1082	0.0030 (-36.19*)
JM - S	-0.1658	0.0023 (-71.94*)	JM - S	-0.1096	0.0029 (-37.30*)
JM - LW	-0.0333	0.0045 (-7.45*)	JM - LW	-0.0191	0.0052 (-3.69*)
V - H	-0.1671	0.0073 (-22.87*)	V - H	-0.1050	0.0088 (-11.90*)
V - S	-0.1738	0.0091 (-19.20*)	V - S	-0.1064	0.0115 (-9.26*)
V - LW	-0.0412	0.0072 (-5.71*)	V - LW	-0.0159	0.0078 (-2.04*)
V - JM	-0.0080	0.0088 (-0.90)	V - JM	0.0032	0.0100 (0.32)
SV - H	-0.1690	0.0074 (-22.96*)	SV - H	-0.1063	0.0088 (-12.02*)
SV - S	-0.1757	0.0091 (-19.32*)	SV - S	-0.1076	0.0115 (-9.37*)
SV - LW	-0.0432	0.0073 (-5.94*)	SV - LW	-0.0171	0.0078 (-2.21*)
SV - JM	-0.0099	0.0089 (-1.11)	SV - JM	0.0019	0.0101 (0.19)
SV - V	-0.0019	0.0003 (-6.72*)	SV - V	-0.0013	0.0003 (-4.42*)
SVlog- H	-0.1696	0.0074 (-23.04*)	SVlog- H	-0.1067	0.0088 (-12.10*)
SVlog- S	-0.1763	0.0091 (-19.42*)	SVlog- S	-0.1081	0.0115 (-9.43*)
SVlog- LW	-0.0438	0.0073 (-6.02*)	SVlog- LW	-0.0176	0.0078 (-2.26*)
SVlog- JM	-0.0105	0.0089 (-1.18)	SVlog- JM	0.0015	0.0100 (0.15)
SVlog- V	-0.0025	0.0003 (-8.24*)	SVlog- V	-0.0017	0.0003 (-5.70*)
SVlog- SV	-0.0006	0.0001 (-8.16*)	SVlog- SV	-0.0004	0.0001 (-6.82*)
TV - H	-0.1764	0.0042 (-41.54*)	TV - H	-0.1230	0.0039 (-31.78*)
TV - S	-0.1831	0.0050 (-36.85*)	TV - S	-0.1244	0.0054 (-23.14*)
TV - LW	-0.0506	0.0049 (-10.24*)	TV - LW	-0.0339	0.0047 (-7.28*)
TV - JM	-0.0173	0.0044 (-3.95*)	TV - JM	-0.0149	0.0037 (-4.01*)
TV - V	-0.0094	0.0092 (-1.02)	TV - V	-0.0180	0.0096 (-1.88)
TV - SV	-0.0075	0.0093 (-0.80)	TV - SV	-0.0168	0.0096 (-1.74)
TV - SVlog	-0.0069	0.0093 (-0.74)	TV - SVlog	-0.0164	0.0096 (-1.70)

Table 6.2: **Eigen-distance and eigen-distance differences (T=200 weeks)** This table reports the 100 times simulation results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 200 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
H	2.8362	0.0093	H	4.1561	0.0137
S	2.2948	0.0033	S	3.6952	0.0108
LW	1.9861	0.0075	LW	3.5318	0.0127
SVlog	1.6868	0.0070	V	3.2984	0.0104
TV	1.6866	0.0060	TV	3.2964	0.0105
V	1.6850	0.0064	SV	3.2958	0.0104
SV	1.6849	0.0069	SVlog	3.2956	0.0104
JM	1.6431	0.0062	JM	3.2778	0.0107
Panel B: Eigen-distance Differences					
	mean	se	t-stats	mean	se	t-stats
S - H	-0.5415	0.0100	(-54.38*)	S - H	-0.4610	0.0140	(-32.82*)
LW - H	-0.8502	0.0090	(-94.12*)	LW - H	-0.6243	0.0090	(-69.43*)
LW - S	-0.3087	0.0075	(-41.15*)	LW - S	-0.1633	0.0121	(-13.54*)
JM - H	-1.1932	0.0100	(-119.28*)	JM - H	-0.8783	0.0100	(-88.20*)
JM - S	-0.6517	0.0062	(-104.74*)	JM - S	-0.4174	0.0089	(-46.91*)
JM - LW	-0.3430	0.0073	(-46.91*)	JM - LW	-0.2540	0.0063	(-40.37*)
V - H	-1.1514	0.0114	(-100.61*)	V - H	-0.8578	0.0141	(-60.87*)
V - S	-0.6099	0.0074	(-82.07*)	V - S	-0.3968	0.0091	(-43.75*)
V - LW	-0.3012	0.0095	(-31.74*)	V - LW	-0.2335	0.0115	(-20.37*)
V - JM	0.0418	0.0093	(4.48*)	V - JM	0.0206	0.0092	(2.24*)
SV - H	-1.1514	0.0117	(-98.49*)	SV - H	-0.8603	0.0140	(-61.40*)
SV - S	-0.6099	0.0078	(-78.30*)	SV - S	-0.3994	0.0091	(-44.02*)
SV - LW	-0.3012	0.0099	(-30.40*)	SV - LW	-0.2360	0.0117	(-20.10*)
SV - JM	0.0418	0.0097	(4.30*)	SV - JM	0.0180	0.0093	(1.94)
SV - V	0.0000	0.0025	(-0.01)	SV - V	-0.0026	0.0022	(-1.16)
SVlog- H	-1.1494	0.0118	(-97.71*)	SVlog- H	-0.8605	0.0140	(-61.36*)
SVlog- S	-0.6079	0.0078	(-77.52*)	SVlog- S	-0.3995	0.0091	(-43.88*)
SVlog- LW	-0.2992	0.0099	(-30.09*)	SVlog- LW	-0.2362	0.0117	(-20.12*)
SVlog- JM	0.0438	0.0098	(4.47*)	SVlog- JM	0.0178	0.0093	(1.92)
SVlog- V	0.0020	0.0026	(0.75)	SVlog- V	-0.0027	0.0022	(-1.25)
SVlog- SV	0.0020	0.0004	(4.82*)	SVlog- SV	-0.0002	0.0003	(-0.74)
TV - H	-1.1496	0.0117	(-98.00*)	TV - H	-0.8598	0.0140	(-61.32*)
TV - S	-0.6081	0.0069	(-88.41*)	TV - S	-0.3988	0.0083	(-47.83*)
TV - LW	-0.2994	0.0097	(-30.89*)	TV - LW	-0.2355	0.0112	(-20.97*)
TV - JM	0.0436	0.0084	(5.19*)	TV - JM	0.0185	0.0089	(2.08*)
TV - V	0.0018	0.0059	(0.30)	TV - V	-0.0020	0.0040	(-0.51)
TV - SV	0.0018	0.0065	(0.28)	TV - SV	0.0005	0.0047	(0.12)
TV - SVlog	-0.0002	0.0066	(-0.03)	TV - SVlog	0.0007	0.0047	(0.16)

Table 6.3: **RMSE and RMSE differences (T=600 weeks)** This table reports the 100 times simulation results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 600 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
S	0.9149	0.0056	S	1.1025	0.0061
JM	0.6317	0.0062	JM	0.8793	0.0062
H	0.6274	0.0054	H	0.8733	0.0056
LW	0.5982	0.0055	LW	0.8509	0.0058
TV	0.5839	0.0065	TV	0.8437	0.0064
SV	0.5778	0.0055	SV	0.8358	0.0064
SVlog	0.5767	0.0055	SVlog	0.8351	0.0064
V	0.5747	0.0055	V	0.8336	0.0064
Panel B: RMSE Differences					
	mean	se	t-stats	mean	se	t-stats
S - H	0.2875	0.0033	(87.57*)	S - H	0.2292	0.0042	(54.97*)
LW - H	-0.0292	0.0015	(-19.58*)	LW - H	-0.0224	0.0014	(-15.72*)
LW - S	-0.3167	0.0039	(-80.54*)	LW - S	-0.2516	0.0043	(-59.08*)
JM - H	0.0043	0.0024	(1.81)	JM - H	0.0060	0.0026	(2.35*)
JM - S	-0.2832	0.0018	(-156.76*)	JM - S	-0.2232	0.0021	(-108.88*)
JM - LW	0.0335	0.0030	(11.31*)	JM - LW	0.0284	0.0027	(10.48*)
V - H	-0.0527	0.0035	(-14.88*)	V - H	-0.0397	0.0048	(-8.31*)
V - S	-0.3402	0.0046	(-74.75*)	V - S	-0.2689	0.0064	(-41.96*)
V - LW	-0.0235	0.0036	(-6.56*)	V - LW	-0.0173	0.0045	(-3.81*)
V - JM	-0.0570	0.0042	(-13.57*)	V - JM	-0.0457	0.0056	(-8.21*)
SV - H	-0.0496	0.0036	(-13.86*)	SV - H	-0.0375	0.0048	(-7.87*)
SV - S	-0.3371	0.0046	(-73.85*)	SV - S	-0.2667	0.0064	(-41.85*)
SV - LW	-0.0204	0.0036	(-5.62*)	SV - LW	-0.0151	0.0045	(-3.32*)
SV - JM	-0.0539	0.0042	(-12.73*)	SV - JM	-0.0435	0.0055	(-7.85*)
SV - V	0.0031	0.0002	(12.53*)	SV - V	0.0022	0.0003	(7.75*)
SVlog- H	-0.0507	0.0036	(-14.15*)	SVlog- H	-0.0382	0.0048	(-8.02*)
SVlog- S	-0.3382	0.0046	(-74.13*)	SVlog- S	-0.2674	0.0064	(-41.98*)
SVlog- LW	-0.0215	0.0036	(-5.90*)	SVlog- LW	-0.0158	0.0045	(-3.48*)
SVlog- JM	-0.0550	0.0042	(-13.00*)	SVlog- JM	-0.0442	0.0055	(-7.98*)
SVlog- V	0.0020	0.0003	(7.45*)	SVlog- V	0.0015	0.0003	(5.13*)
SVlog- SV	-0.0011	0.0001	(-19.23*)	SVlog- SV	-0.0007	0.0000	(-16.88*)
TV - H	-0.0435	0.0027	(-16.26*)	TV - H	-0.0296	0.0024	(-12.12*)
TV - S	-0.3310	0.0039	(-84.29*)	TV - S	-0.2588	0.0043	(-60.27*)
TV - LW	-0.0143	0.0028	(-5.03*)	TV - LW	-0.0072	0.0025	(-2.92*)
TV - JM	-0.0478	0.0031	(-15.51*)	TV - JM	-0.0356	0.0029	(-12.11*)
TV - V	0.0092	0.0047	(1.97*)	TV - V	0.0101	0.0055	(1.83)
TV - SV	0.0061	0.0047	(1.30)	TV - SV	0.0079	0.0055	(1.43)
TV - SVlog	0.0072	0.0047	(1.54)	TV - SVlog	0.0086	0.0055	(1.56)

Table 6.4: **Eigen-distance and eigen-distance differences (T=600 weeks)** This table reports the 100 times simulation results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 600 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
S	2.2599	0.0020	S	2.6676	0.0065
H	1.4596	0.0039	H	2.0838	0.0062
LW	1.3032	0.0038	LW	1.9714	0.0059
JM	1.2263	0.0039	JM	1.8691	0.0050
SV	1.1533	0.0030	SV	1.8496	0.0048
SVlog	1.1523	0.0031	SVlog	1.8488	0.0047
TV	1.1456	0.0037	V	1.8440	0.0045
V	1.1401	0.0031	TV	1.8431	0.0050
Panel B: Eigen-distance Differences					
	mean	se t-stats		mean	se t-stats
S - H	0.8003	0.0044 (182.62*)	S - H	0.5839	0.0079 (73.67*)
LW - H	-0.1564	0.0018 (-87.20*)	LW - H	-0.1124	0.0017 (-64.70*)
LW - S	-0.9566	0.0042 (-228.58*)	LW - S	-0.6962	0.0078 (-89.10*)
JM - H	-0.2333	0.0052 (-44.74*)	JM - H	-0.2147	0.0059 (-36.31*)
JM - S	-1.0336	0.0035 (-296.87*)	JM - S	-0.7986	0.0053 (-150.22*)
JM - LW	-0.0769	0.0050 (-15.27*)	JM - LW	-0.1023	0.0056 (-18.19*)
V - H	-0.3195	0.0056 (-56.95*)	V - H	-0.2397	0.0075 (-32.02*)
V - S	-1.1198	0.0039 (-290.10*)	V - S	-0.8236	0.0073 (-112.35*)
V - LW	-0.1631	0.0053 (-30.93*)	V - LW	-0.1274	0.0071 (-17.87*)
V - JM	-0.0862	0.0050 (-17.37*)	V - JM	-0.0250	0.0057 (-4.37*)
SV - H	-0.3063	0.0055 (-55.59*)	SV - H	-0.2342	0.0077 (-30.61*)
SV - S	-1.1065	0.0035 (-314.12*)	SV - S	-0.8181	0.0074 (-111.18*)
SV - LW	-0.1499	0.0052 (-28.85*)	SV - LW	-0.1218	0.0073 (-16.75*)
SV - JM	-0.0730	0.0047 (-15.45*)	SV - JM	-0.0195	0.0059 (-3.30*)
SV - V	0.0132	0.0018 (7.44*)	SV - V	0.0055	0.0016 (3.46*)
SVlog- H	-0.3073	0.0055 (-55.55*)	SVlog- H	-0.2350	0.0076 (-30.80*)
SVlog- S	-1.1076	0.0036 (-310.60*)	SVlog- S	-0.8189	0.0073 (-111.47*)
SVlog- LW	-0.1509	0.0052 (-28.98*)	SVlog- LW	-0.1226	0.0072 (-16.93*)
SVlog- JM	-0.0740	0.0048 (-15.53*)	SVlog- JM	-0.0203	0.0059 (-3.44*)
SVlog- V	0.0122	0.0018 (6.69*)	SVlog- V	0.0047	0.0016 (2.96*)
SVlog- SV	-0.0010	0.0003 (-3.52*)	SVlog- SV	-0.0008	0.0002 (-3.49*)
TV - H	-0.3140	0.0059 (-52.97*)	TV - H	-0.2407	0.0075 (-32.17*)
TV - S	-1.1143	0.0044 (-254.46*)	TV - S	-0.8245	0.0078 (-106.18*)
TV - LW	-0.1577	0.0057 (-27.60*)	TV - LW	-0.1283	0.0071 (-18.00*)
TV - JM	-0.0807	0.0054 (-15.07*)	TV - JM	-0.0260	0.0059 (-4.41*)
TV - V	0.0055	0.0025 (2.23*)	TV - V	-0.0009	0.0023 (-0.40)
TV - SV	-0.0077	0.0028 (-2.74*)	TV - SV	-0.0065	0.0028 (-2.29*)
TV - SVlog	-0.0067	0.0028 (-2.37*)	TV - SVlog	-0.0057	0.0028 (-2.01*)

Table 6.5: **RMSE and RMSE differences (T=1000 weeks)** This table reports the 100 times simulation results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 1000 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
S	0.8756	0.0040	S	1.0024	0.0067
JM	0.5540	0.0046	JM	0.7373	0.0070
H	0.4857	0.0043	H	0.6869	0.0063
LW	0.4728	0.0043	LW	0.6776	0.0065
SV	0.4643	0.0044	SV	0.6743	0.0067
SVlog	0.4631	0.0044	SVlog	0.6734	0.0067
TV	0.4598	0.0045	V	0.6705	0.0067
V	0.4585	0.0045	TV	0.6702	0.0067
Panel B: RMSE Differences					
	mean	se	t-stats	mean	se	t-stats
S - H	0.3899	0.0033	(117.68*)	S - H	0.3154	0.0041	(76.33*)
LW - H	-0.0129	0.0009	(-15.10*)	LW - H	-0.0093	0.0009	(-10.49*)
LW - S	-0.4028	0.0037	(-110.09*)	LW - S	-0.3247	0.0046	(-71.37*)
JM - H	0.0683	0.0025	(27.42*)	JM - H	0.0504	0.0028	(18.17*)
JM - S	-0.3216	0.0016	(-204.35*)	JM - S	-0.2651	0.0018	(-146.26*)
JM - LW	0.0812	0.0029	(28.20*)	JM - LW	0.0597	0.0032	(18.79*)
V - H	-0.0272	0.0028	(-9.83*)	V - H	-0.0164	0.0034	(-4.77*)
V - S	-0.4171	0.0040	(-105.26*)	V - S	-0.3318	0.0056	(-59.49*)
V - LW	-0.0143	0.0029	(-5.00*)	V - LW	-0.0071	0.0035	(-2.04*)
V - JM	-0.0955	0.0035	(-27.29*)	V - JM	-0.0668	0.0047	(-14.28*)
SV - H	-0.0214	0.0028	(-7.78*)	SV - H	-0.0126	0.0034	(-3.70*)
SV - S	-0.4113	0.0040	(-104.02*)	SV - S	-0.3281	0.0056	(-59.00*)
SV - LW	-0.0085	0.0029	(-2.99*)	SV - LW	-0.0034	0.0035	(-0.97)
SV - JM	-0.0897	0.0035	(-25.62*)	SV - JM	-0.0630	0.0047	(-13.53*)
SV - V	0.0058	0.0002	(27.71*)	SV - V	0.0037	0.0002	(15.58*)
SVlog- H	-0.0226	0.0028	(-8.20*)	SVlog- H	-0.0135	0.0034	(-3.94*)
SVlog- S	-0.4125	0.0039	(-104.51*)	SVlog- S	-0.3289	0.0056	(-59.24*)
SVlog- LW	-0.0097	0.0029	(-3.39*)	SVlog- LW	-0.0042	0.0035	(-1.21)
SVlog- JM	-0.0909	0.0035	(-26.01*)	SVlog- JM	-0.0639	0.0047	(-13.73*)
SVlog- V	0.0046	0.0002	(21.11*)	SVlog- V	0.0029	0.0002	(12.07*)
SVlog- SV	-0.0012	0.0000	(-26.35*)	SVlog- SV	-0.0008	0.0000	(-21.20*)
TV - H	-0.0259	0.0014	(-19.16*)	TV - H	-0.0167	0.0013	(-12.70*)
TV - S	-0.4158	0.0032	(-128.93*)	TV - S	-0.3321	0.0041	(-81.80*)
TV - LW	-0.0130	0.0017	(-7.83*)	TV - LW	-0.0074	0.0017	(-4.46*)
TV - JM	-0.0942	0.0024	(-39.18*)	TV - JM	-0.0671	0.0027	(-24.97*)
TV - V	0.0013	0.0030	(0.45)	TV - V	-0.0003	0.0036	(-0.09)
TV - SV	-0.0045	0.0030	(-1.52)	TV - SV	-0.0040	0.0036	(-1.13)
TV - SVlog	-0.0033	0.0030	(-1.11)	TV - SVlog	-0.0032	0.0036	(-0.90)

Table 6.6: **Eigen-distance and eigen-distance differences (T=1000 weeks)** This table reports the 100 times simulation results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 1000 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
S	2.2499	0.0015	S	2.4823	0.0056
JM	1.1461	0.0027	H	1.5629	0.0035
H	1.1062	0.0030	JM	1.5508	0.0042
LW	1.0346	0.0028	LW	1.5114	0.0034
SV	0.9504	0.0027	SV	1.4528	0.0038
SVlog	0.9489	0.0027	SVlog	1.4521	0.0038
V	0.9328	0.0024	V	1.4489	0.0039
TV	0.9326	0.0024	TV	1.4476	0.0040
Panel B: Eigen-distance Differences					
	mean	se t-stats		mean	se t-stats
S - H	1.1437	0.0034 (331.77*)	S - H	0.9194	0.0065 (141.22*)
LW - H	-0.0716	0.0009 (-81.36*)	LW - H	-0.0515	0.0010 (-53.57*)
LW - S	-1.2153	0.0032 (-376.11*)	LW - S	-0.9709	0.0064 (-151.48*)
JM - H	0.0399	0.0040 (9.86*)	JM - H	-0.0121	0.0047 (-2.57*)
JM - S	-1.1038	0.0027 (-405.64*)	JM - S	-0.9316	0.0047 (-198.26*)
JM - LW	0.1115	0.0038 (29.54*)	JM - LW	0.0393	0.0044 (8.87*)
V - H	-0.1734	0.0038 (-45.38*)	V - H	-0.1140	0.0048 (-23.80*)
V - S	-1.3171	0.0029 (-458.31*)	V - S	-1.0334	0.0070 (-148.15*)
V - LW	-0.1018	0.0036 (-28.07*)	V - LW	-0.0625	0.0047 (-13.33*)
V - JM	-0.2133	0.0036 (-58.74*)	V - JM	-0.1019	0.0052 (-19.73*)
SV - H	-0.1559	0.0041 (-38.31*)	SV - H	-0.1101	0.0048 (-22.93*)
SV - S	-1.2995	0.0033 (-397.72*)	SV - S	-1.0295	0.0069 (-148.66*)
SV - LW	-0.0842	0.0039 (-21.76*)	SV - LW	-0.0586	0.0047 (-12.37*)
SV - JM	-0.1957	0.0038 (-51.34*)	SV - JM	-0.0980	0.0053 (-18.51*)
SV - V	0.0176	0.0016 (10.93*)	SV - V	0.0039	0.0016 (2.39*)
SVlog- H	-0.1574	0.0041 (-38.72*)	SVlog- H	-0.1108	0.0048 (-23.10*)
SVlog- S	-1.3010	0.0033 (-397.11*)	SVlog- S	-1.0303	0.0069 (-149.30*)
SVlog- LW	-0.0857	0.0039 (-22.20*)	SVlog- LW	-0.0594	0.0047 (-12.56*)
SVlog- JM	-0.1972	0.0038 (-51.88*)	SVlog- JM	-0.0987	0.0053 (-18.78*)
SVlog- V	0.0160	0.0016 (9.83*)	SVlog- V	0.0032	0.0016 (2.00*)
SVlog- SV	-0.0015	0.0003 (-5.16*)	SVlog- SV	-0.0007	0.0002 (-3.44*)
TV - H	-0.1736	0.0040 (-43.21*)	TV - H	-0.1153	0.0048 (-24.15*)
TV - S	-1.3173	0.0028 (-462.27*)	TV - S	-1.0347	0.0069 (-149.40*)
TV - LW	-0.1020	0.0038 (-26.84*)	TV - LW	-0.0638	0.0047 (-13.68*)
TV - JM	-0.2135	0.0037 (-58.25*)	TV - JM	-0.1031	0.0052 (-19.99*)
TV - V	-0.0002	0.0012 (-0.18)	TV - V	-0.0013	0.0012 (-1.01)
TV - SV	-0.0178	0.0021 (-8.50*)	TV - SV	-0.0052	0.0020 (-2.57*)
TV - SVlog	-0.0163	0.0021 (-7.73*)	TV - SVlog	-0.0044	0.0020 (-2.26*)

Chapter 7

Extensions of simulation analysis to more realistic covariance structures

7.1 Introduction

So far our simulations have been based on the assumption of a constant covariance structure for stock returns following multivariate normal distributions. However, it is widely agreed now that in reality financial asset returns have fat-tailed distributions and their volatilities and correlations are time varying and often exhibit some long range dependence.[1]

This chapter extends our previous analysis by using more realistic assumptions for a return covariance structure to incorporate some of the properties observed in the

[1]See for example, Campbell (1996), Ferson and Harvey (1991), Jagannathan and Wang (1996) and Campbell, Lettau, Malkiel and Xu (2001).

empirical covariance matrices. We consider two extensions. The first extension considers a covariance structure of the returns that have fat-tailed distributions, but we still assume that the covariance structure is constant. The second extension considers a more interesting and economically motivated time-varying covariance structure where stocks migrate among different risk categories. We study how the relative performance of the various covariance estimators we have studied in Chapter 6 is affected by these more realistic covariance structures and what types of methods become more important under these circumstances. In both extensions, we find that our eigenvalue shrinkage estimators are still very useful if not become more important.

For the covariance structure of returns with excess kurtosis (fat-tails) in the distributions, we use a multivariate normal inverse Gaussian (MNIG) distribution to model the log-returns of asset prices. This family of distributions has proven to fit the heavy tails observed in financial time series extremely well.

For the time-varying covariance structure, we use a mean-reverting Ornstein-Uhlenbeck (OU) process to model a particular time-varying covariance structure, where the cross section of risk characteristics tends not to change but the identities of stocks do. In other words, the general risk characteristics of the market remain stationary even though the individual stocks migrate among risk classes. We assume the return generating factors are constant but the factor loadings follow a mean-reverting OU process. This is inspired by the findings that the estimated betas in the Sharpe

single-factor model exhibit a tendency to regress over time towards the grand mean of all betas (Blume (1971)).

The rest of this chapter is organized as follows. Section 7.2 is the extension of the analysis when the true covariance structure of the returns contain excess Kurtosis. We review a particular multivariate normal inverse Gaussian (MNIG) distribution and show how we use it to construct our covariance structure. We then analyze the performance of different covariance estimators under this true covariance structure following the comparison methods we have used in Chapter 6. Section 7.3 extends the analysis to the time-varying covariance structure. We review a general Ornstein-Uhlenbeck (OU) process and specify a time-varying covariance structure for our numerical experiment. We then discuss the performance of the alternative covariance estimators under this time-varying covariance structure. Section 7.4 concludes.

7.2 Extension to a covariance structure for returns with fat-tailed distributions

7.2.1 Multivariate normal inverse Gaussian distribution

The empirical returns of stocks are found to have fat-tailed distributions.[2] There has been an increasing interest in using a normal inverse Gaussian (NIG) distribution to

[2]For example, for the empirical data we have used in Chapter 3, we find that the average kurtosis of the 78 stocks equals 4.60, 4.33 and 4.78 for the first sample period, second sample period and combined first and second sample periods respectively. The kurtosis for the normal distribution is 3. Distributions that have kurtosis greater than 3 have fat tails.

model the heavy-tailed stock returns. A NIG distribution, NIG($\alpha, \beta, \mu, \delta$), is a closed form distribution that is completely specified by its four real valued parameters, which have natural interpretations in terms of shapes of the distribution. α controls the steepness of the density, and large values of α imply light tails and vice versa. β is a vector skewness parameter, δ is a scalar parameter and μ is a vector translation parameter. The multivariate normal inverse Gaussian, MNIG($\alpha, \beta, \mu, \delta, \mathbf{\Gamma}$), is the generalization of a multivariate Gaussian distribution with a NIG mixing distribution. $\mathbf{\Gamma}$ controls the degree of correlation between the heavy-tailed components.[3]

7.2.2 Specifications for MNIG simulations

We follow the work by Oigard, Hanssen and Hansen (2004) to construct the stock returns, $\mathbf{X}$, which are assumed to follow an MNIG($\alpha, \beta, \mu, \delta, \mathbf{\Gamma}$) distribution using[4]

$$\mathbf{X} = \mu + Z\mathbf{\Gamma}\beta + \sqrt{Z}\mathbf{\Gamma}^{1/2}\mathbf{Y} \tag{7.2.1}$$

where $Z \sim \text{IG}(\delta^2, \alpha^2 - \beta^T\mathbf{\Gamma}\beta)$ where IG(χ, ψ) denotes the inverse Gaussian (IG) distribution, and $\mathbf{Y} \sim \text{N}(\mathbf{0}, \mathbf{I})$. The mean vector and the covariance matrix of $\mathbf{X}$ are obtained as

$$E(\mathbf{X}) = \mu + \frac{\delta\mathbf{\Gamma}\beta}{\sqrt{\alpha^2 - \beta^T\mathbf{\Gamma}\beta}} \tag{7.2.2}$$

$$\mathbf{\Sigma} = \delta(\alpha^2 - \beta^T\mathbf{\Gamma}\beta)^{-1/2}\left[\mathbf{\Gamma} + (\alpha^2 - \beta^T\mathbf{\Gamma}\beta)^{-1}\mathbf{\Gamma}\beta\beta^T\mathbf{\Gamma}\right] \tag{7.2.3}$$

---

[3] See for example, Barndorff-Nielsen (1997a, 1997b), Barndorff-Nielsen and Shephard (2001), Rydberg (1997), Eberlein and Keller (1995), Oigard, Hanssen and Hansen (2004), and Benth, Groth and Kettler (2006).

[4] α and δ are scalar parameters, β and μ are vector parameters, and $\mathbf{\Gamma}$ is a matrix parameter, and $\alpha > 0$, $\beta \in \mathbb{R}^d$, $\delta > 0$, $\mu \in \mathbb{R}^d$, and $\mathbf{\Gamma} \in \mathbb{R}^{d \times d}$.

We assume the following parameter values: $\alpha = 0.5$, $\beta = 0$, $\mu = 0$, $\delta = 0.5$, so that the returns have zero means, variances equal to one, zero skewness and kurtosis equal to 15. The mean and covariance matrix of $\mathbf{X}$ are then simplified and equal to

$$E(\mathbf{X}) = \mathbf{0} \tag{7.2.4}$$

$$\mathbf{\Sigma} = \frac{\delta}{\alpha}\mathbf{\Gamma} \tag{7.2.5}$$

We also assume that $\mathbf{\Sigma} = \mathbf{C}_{true}$. $\mathbf{X}$ can therefore be simulated from:

$$\mathbf{X} = \sqrt{Z}\left(\frac{\alpha}{\delta}\mathbf{C}_{true}\right)^{1/2}\mathbf{Y} \tag{7.2.6}$$

We then use the quasi-Monte Carlo method proposed by Benth, Groth and Kettler (2006) to simulate the IG distributed random variable Z ($Z \sim \text{IG}(\delta^2, \alpha^2)$). Under their method, the sampling of Z consists of firstly drawing a random variable W from V, which is χ^2_1-distributed

$$W = \xi + \frac{\xi^2 V}{2\delta^2} - \frac{\xi}{2\delta^2}\sqrt{4\xi\delta^2 V + \xi^2 V^2} \tag{7.2.7}$$

and then letting

$$Z = W \cdot \mathbf{1}_{\left(U_1 \leq \frac{\xi}{\xi+W}\right)} + \frac{\xi^2}{W} \cdot \mathbf{1}_{\left(U_1 \geq \frac{\xi}{\xi+W}\right)} \tag{7.2.8}$$

where U_1 is uniformly distributed and $\xi = \delta/\sqrt{\alpha^2 - \beta^2}$.

7.2.3 Results

Table 7.1 - Table 7.6 report the average RMSE and eigen-distance (Panel A) and their differences (Panel B) for the covariance structure of returns with fat-tail distributions

when the size of time-series observations equals 200, 600 and 1,000 weeks respectively.

We find that overall the average values of RMSE and eigen-distance results of different estimators are larger with bigger standard errors and a lot closer to each other than those under the corresponding constant covariance structure.

We also find that the results for $T = 1000$ under the MNIG scenario are similar to those for $T = 600$ under the covariance structure for multivariate normal returns, both in terms of goodness of estimates and best estimators. More specifically, the results of Table 7.5 and Table 7.6 (for MNIG T=1000) are similar to those of Table 6.3 and Table 6.4 (for T=600) in Chapter 6.

In other words, the improvement of the performance of our shrinkage estimators relative to the Jagannathan and Ma estimator is a lot slower when they are under the covariance structure of the returns with fat-tailed distributions than when they are under the constant covariance structure of returns with normal distributions. It takes a larger sample of observations to out-perform the Jagannathan and Ma estimator. For example, when the comparisons are made against the true covariance structure using the eigen-distance measure, our four shrinkage estimators out-perform the Jagannathan and Ma estimator when T=600 under the multivariate normal covariance structure. But under the MNIG scenario, our four shrinkage estimators out-perform the Jagannathan and Ma estimator only when T=1000. Nevertheless, we still find

that our simpler shrinkage estimators are consistently better than the more complicated Ledoit and Wolf shrinkage estimator.[5]

7.3 Extension to a time-varying covariance structure

A conventional method to characterize the time-varying covariance structure of returns is to employ a multivariate GARCH model. The GARCH model and their extensions have been regarded as good econometric models to capture certain facts of the empirical volatility process. We have reviewed this approach briefly in Chapter 2 and noted some of the difficulties involved. As usual, this type of models also pose a dilemma concerning a choice between insufficient structure and too many parameters.

In this section, we consider a particular time varying structure where the general risk characteristics of the market remain stationary, even though the individual stocks migrate among risk classes. It is not obvious what kind of the multivariate GARCH model would be necessary to have this set of properties. This leads us to develop a different kind of model. We employ a tractable Ornstein-Uhlenbeck (OU) process to model the situation where the cross-section of risk characteristics tends not to change, but the identities of the stocks do.

---

[5]We report that for T=200, 600 and 1000 weeks, the average shrinkage intensity of the Ledoit and Wolf estimator equals 0.34, 0.13 and 0.09, and the average adjustment on the first eigenvector equals 0.85, 0.95 and 0.97 respectively.

7.3.1 A general Ornstein-Uhlenbeck process

The classic Ornstein-Uhlenbeck (OU) process is a unique mean-reverting process which is simultaneously Gaussian, Markov and stationary. More specifically, the OU process is a stochastic process given by the following stochastic differential equation:

$$dX_t = \alpha(\bar{X} - X_t)dt + \sigma dZ_t \tag{7.3.1}$$

where α is the mean reverting speed, $\bar{X}$ is the reverting mean, σ is the noise and $dZ \sim N(0, dt)$. α, $\bar{X}$ and σ are the parameters that define a particular OU process.

Assuming $f(X_t, t) = Xe^{\alpha t}$ and applying Ito's lemma, we can derive the following:

$$E[X_T|X_0] = \bar{X} + (X_0 - \bar{X})e^{-\alpha T} \tag{7.3.2}$$

$$Var[X_T|X_0] = \sigma^2 \int_0^T e^{-2\alpha(T-t)} = \frac{\sigma^2}{2\alpha}[1 - e^{-2\alpha T}] \tag{7.3.3}$$

where X_T is a realization of the O-U process at time T. The asymptotic values of the above results in the limiting case equal:

$$\lim_{T\to\infty} E[X_T|X_0] = \bar{X} \tag{7.3.4}$$

$$\lim_{T\to\infty} Var[X_T|X_0] = \frac{\sigma^2}{2\alpha} \tag{7.3.5}$$

Using Equation (7.3.5), we can obtain the asymptotic standard deviation $SD(X)$, as:

$$SD(X) = \lim_{T\to\infty} SD(X_T) = \frac{\sigma}{\sqrt{2\alpha}} \tag{7.3.6}$$

The noise term σ therefore equals

$$\sigma = SD(X) \times \sqrt{2\alpha} \tag{7.3.7}$$

This equation is useful since if we can estimate $SD(X)$ and either α or σ, it enables us to obtain the other parameter.

7.3.2 Simulate returns from an N-variate factor model with factor loadings following Ornstein-Uhlenbeck processes

Recall that in Chapter 4, where the true covariance structure is assumed to be constant, the stock return r_i is simulated using a N-variate factor model where the kth-factor is normally distributed with zero mean and variance equals the kth time-adjusted eigenvalue λ_k^*, and the kth-factor loading equals the ith element of the eigenvector t_k.[6]

In the case of a mean-reverting time-varying covariance structure, we assume that the N factors are still multivariately normally distributed with zero means and variances equal the adjusted eigenvalues, but the factor loadings on each stock i, $t_{i,k}$

[6]Using the notations in Chapter 4, we have

$$\begin{aligned} f_k &\propto \mathbf{N}(0, \lambda_k^*), k = 1, \cdots, N \\ b_{k,i} &= t_{i,k} \\ r_i &= \sum_{k=1}^{N} t_{i,k} f_k \end{aligned}$$

where r_i is the simulated return for stock i, f_k is the kth factor, and $b_{k,i}$ is the loading for the kth factor on each stock i. λ_k^* is the kth eigenvalue of the true covariance matrix $\mathbf{C_{true}}$, $t_{i,k}$ corresponds to the element of the ith row of the kth eigenvector of the same true covariance matrix. For more details, refer to Section 4.3 in Chapter 4.

(for each factor k ($k = 1, \ldots, N$)), follow processes based on the mean-reverting O-U processes,

$$dt_{i,k} = \alpha_k(\bar{t}_k - t_{i,k})dt + \sigma_k dZ \tag{7.3.8}$$

where α_k is the reverting parameter and σ_k is the noise contributor. $\bar{t}_k$ is the reverting mean or the cross-sectional average of the factor loadings of all stocks ($i = 1, \ldots, N$), and we assume $\bar{t}_k = \frac{1}{N}\sum_{i=1}^{i=N} t_{i,k}$. We will describe later how we re-orthogonalize and normalize these vectors in order to preserve the overall risk structure. We simulate weekly data using $\delta t = \frac{1}{52}$ and $\delta Z = \sqrt{\frac{1}{52}}W(W \sim N(0,1))$ (assuming 52 weeks per annum).

To simulate the time-varying factor loadings, we need to define the parameter values α_k, σ_k and $\bar{t}_k$. Since we know $\bar{t}_k$, we can calculate the volatility of the cross-sectional average factor loading across stocks as

$$SD(t_k) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{i=N}(t_{i,k} - \bar{t}_k)^2} \tag{7.3.9}$$

According to Equation (7.3.5), if we specify a value of α_k, we can then obtain σ_k as:

$$\sigma_k = SD(t_k) \times \sqrt{2\alpha_k} \tag{7.3.10}$$

Thus the problem of determining the Ornstein-Uhlenbeck process for the factor loadings becomes the issue of specifying the value of the reverting parameter α_k.

Blume (1971) finds that estimated betas exhibit a tendency to regress over time towards the grand mean of all betas. In his subsequent paper, Blume (1975) derives

a method[7] to separate the estimation error (Blume calls it *order-bias*) from the true regression and he finds that on average stocks regress towards their mean about 20% over seven years (or roughly 3% a year).

We use the Blume (1975) results to help calibrate the mean-reverting parameters. We assume that the 3% per year mean reverting speed that Blume finds for betas applies to the loadings of all factors, i.e., $\alpha_k = 0.03$ ($k \in [1, N]$).

7.3.3 Estimating covariance matrices under a time-varying covariance structure

The time-varying covariance structure raises some new issues in addition to those we have already been dealing with under a constant covariance structure. This section describes the problems we encounter when we consider the particular type of the time-varying covariance structure we have described in the previous section, and the approaches we take to tackle these problems.

There are three main problems: (1) how to preserve the appropriate eigenvalue structure. We have been constructing a true covariance matrix and stock returns from

[7] Blume (1975) adjusts the estimated betas using the following formula

$$E(\beta_{it}|\hat{\beta}_{it}) - 1 = \frac{\sigma^2(\beta_{it})}{\sigma^2(\hat{\beta}_{it})}(\hat{\beta}_{it} - 1) \tag{7.3.11}$$

where $\sigma^2(\hat{\beta}_{it})$ is the cross-sectional variance of the estimated betas of all stocks in the sample. The estimate of $\sigma^2(\beta_{it})$ is derived as the difference between the estimates of $\sigma^2(\hat{\beta}_{it})$ and $\sigma^2(\eta_{it})$ ($\sigma^2(\hat{\beta}_{it}) - \sigma^2(\eta_{it})$), where $\sigma^2(\eta_{it})$ is the average of the squares of the standard error associated with each estimated beta (Blume (1975)). Blume shows both theoretically and empirically that a major reason for the observed regression tendency is the real non-stationarities in the underlying values of betas.

a multivariate factor model with the orthogonal eigenvectors and their corresponding eigenvalues. However, since we now assume the eigenvectors follow O-U processes, they are no longer always orthogonal to each other; (2) how to define an appropriate comparison. In the time-varying situation, the true covariance matrix is no longer constant but changes all the time. We need to consider an appropriate benchmark to compare the performance of different estimators; and (3) how to implement certain covariance estimation methods, for example, our eigenvalue shrinkage methods, which involve adjustments over two periods. As the true covariance matrix changes over time, we cannot adjust the eigenvalues of two successive sample covariance matrices (from non-overlapping data) as the earlier one is based on distant information which is out-of-date. The following describes what we have done to solve these problems.

Re-orthogonalization

For each week, we obtain a new set of factor loadings for all the N factors and all the stocks, $\mathbf{T}_t = [\mathbf{t}_{1,t}, \ldots, \mathbf{t}_{N,t}]$, where the subscript t denotes the point of time. As these factor loadings (eigenvectors) change over time, they no longer maintain orthogonality to each other. To solve this problem, we use a standard Gram-Schmidt orthogonalization procedure on these eigenvectors. We then use the adjusted eigenvectors, $\mathbf{T}_t^* = [\mathbf{t}_{1,t}^*, \ldots, \mathbf{t}_{N,t}^*]$, to construct an *instant true* covariance matrix at time t, $\mathbf{C}_{true,t}$, using

$$\mathbf{C}_{true,t} = \mathbf{T}_t^* \mathbf{\Lambda}^* \mathbf{T}_t^{*'} \tag{7.3.12}$$

where $\mathbf{\Lambda}^*$ is the same adjusted eigenvalue matrix we use to construct a constant true covariance matrix.

We then use the N-variate factor model to generate one return for each stock from this *instant true* covariance structure. We repeat this procedure and generate a series of *instant true* covariance matrices over a period of time T, and obtain a time series of stock returns $\mathbf{R}_{ou}$. $\mathbf{R}_{ou}$ is then used to construct the covariance matrices ($\mathbf{\Sigma}_{ou}$'s) according to different estimation methods. $\mathbf{\Sigma}_{ou}$'s are predictions of covariance matrices from the time-series return observations over the period of time T.

Comparison benchmark

As there is no longer a constant *true* covariance matrix to compare with in this case, we compare the covariance estimations $\mathbf{\Sigma}_{ou}$'s with a one-week ahead *instant true* covariance matrix for time $T+1$, $\mathbf{C}_{true,T+1}$, which is constructed as

$$\mathbf{C}_{true,T+1} = \mathbf{T}^*_{T+1}\mathbf{\Lambda}^*\mathbf{T}^{*'}_{T+1} \tag{7.3.13}$$

Note that $\mathbf{T}$ denotes the eigenvector matrix, and T denotes the time.

Eigenvalue adjustment

To adjust the eigenvalues, we simulate stock returns for an arbitrary 200 times from the *instantaneous true* covariance matrix $\mathbf{C}_{true,T+1}$. This gives us the variances and covariances of the returns from T to $T+1$, $\mathbf{C}_{T,T+1}$. We then use $\mathbf{C}_{T,T+1}$ as the second

period covariance matrix to adjust the sample covariance matrix for our eigenvalue shrinkage estimators.[8]

Additional methods to compare with

It seems reasonable to also consider the RiskMetrics and overall mean model in the time-varying extension. This is because the recent information is more relevant now than the more distant information in estimating a time-varying covariance matrix, therefore methods like the RiskMetrics that imposes more weights on recent observations may be more useful. In addition, as the stocks migrate across different risk categories, an average relationship estimator may work better than estimators that try to measure the individual pair-wise relationships.

7.3.4 Results

Table 7.7 - Table 7.12 report the average RMSE and eigen-distance (Panel A) and their differences (Panel B) when the size of time-series observations equals 200, 600 and 1,000 weeks respectively.

First of all, we find that our eigenvalue shrinkage methods (V, $SVlog$ and TV) become more important. In particular, TV is always the best estimator under both measures and all three lengths of time series observations we report. The way we use ex-post information to adjust the eigenvalues has no doubt also contributed to a better

[8]The problem here is that we are using ex-post information to help adjust the eigenvalues. Further studies are needed to explore better ways to adjust eigenvalues.

performance of these eigenvalue shrinkage methods. However, the cross comparison among these methods themselves shows that TV is consistently better than V and SVlog, which shows that adjusting the principal eigenvectors is helpful when the eigenvectors are clearly mean-reverting as in this case we have constructed them to do so. This is opposite to what we find under the constant covariance structure, where we show that if we are already adjusting the eigenvalues, the additional adjustment on the eigenvector does not add much value.[9]

We find that the RiskMetrics estimator, particularly with the decaying factor ($\lambda = 0.95$) we are using, performs very badly by itself. It consistently under-performs the sample historical covariance matrix and all other alternative estimators.[10] However, when we combine the RiskMetrics estimator with a single-index estimator (RM+S) or an overall mean estimator (RM+C), the combined estimator works much better. For example, both (RM+S) and (RM+C) work better than the overall mean estimator (C) under both the RMSE and eigen-distance measures.

When we consider a bigger decaying factor, for example $\lambda = 0.99$, the RiskMetrics performs a lot differently and much better. This shows that a different form of the RiskMetrics estimator, or in general a different estimator that places more weight on the recent observations, may be able to provide a better estimate of the covariance

---

[9]We find that average adjustment on the first eigenvector, $b_{\mathbf{t}_{1,1}}$ in the time-varying case equals to 0.85, 0.91 and 0.90 for T=200, 600 and 1,000 weeks respectively.

[10]We checked the eigenvalues of the RiskMetrics estimated covariance matrix and find the latter part of the eigenvalues is much smaller than those of the other estimated covariance matrices.

structure. At least, we expect the combination of this estimator with a single-index estimator can out-perform the Jagannathan and Ma estimator as it is better to use more recent data in estimating a time-varying covariance structure. It will be an interesting topic for further research.[11]

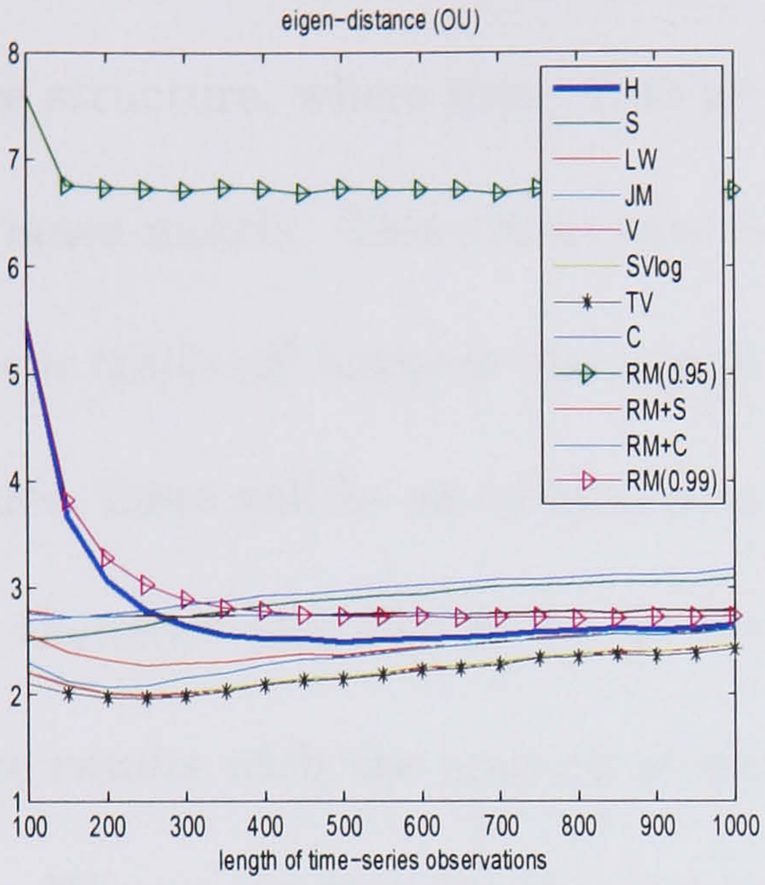
The overall mean estimator does not work well as we expect. This may be because the dynamic migrating structure that we have assumed for the stocks, where their risks due to the N-variate factors change independently and concurrently, is too complicated to be captured by a simple average model. We also suspect that our structure may have a rather slow migration since we expect the average model to perform well asymptotically.[12] This shows that we need more empirical work to choose and calibrate our time-varying model.

The sample covariance matrix performs better with increasing amount of the time series observations. It consistently out-performs the single-index estimator and the overall mean estimator when there are more time series observations used in estimating the covariance. However, the significance of this relative advantage over the single-index estimator is much smaller than that under the constant covariance structure. For example, under the eigen-distance measure when T=600, the t-statistic for $(S-H)$ under the constant covariance structure is 182.6 (Table 6.4), whereas under

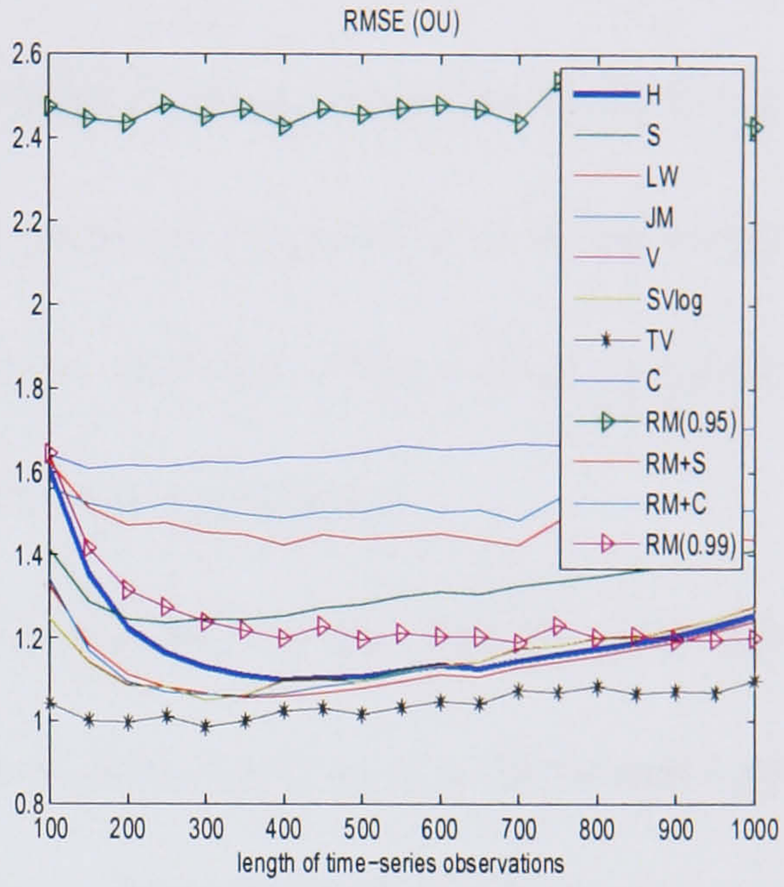
---

[11]We find that the RiskMetrics estimator with λ=0.99 combined with the single-index estimator (RM(0.99)+S) and the overall mean estimator (RM(0.99)+C) (both not reported) did provide significantly better results than (RM(0.95)+S) and (RM(0.95)+C).

[12]We find that the eigen-distance of two instant true covariance matrices $\mathbf{C}_{true,T}$ and $\mathbf{C}_{true,T+1}$ equals 0.1859, 0.2142 and 0.2596 when T=200, 600 and 1000 weeks respectively.



A.



B.

Figure 7.1: **Eigen-distance and RMSE with increasing estimation windows under a time varying structure** This figure shows the eigen-distance and RMSE results of the different covariance estimators when the estimation windows increase from 100 weeks to 1,000 weeks. Figure A (B) plot the eigen-distance (RMSE) results when compared with the *instant true* covariance matrix.

the time-varying covariance structure it is only 39.3 (Table 7.10). This shows that the time variation makes the estimation more difficult to both estimators.

Another estimator whose performance increases with the amount of the observation of is the Ledoit and Wolf estimator. Its relative performance over the Jagannathan and Ma estimator increases when T increases.[13]

We also find that the forecast errors of the individual estimators and the differences between different estimators increase when the amount of time series data used

[13]The average shrinkage intensity for the Ledoit and Wolf estimator in the time varying case equals 0.21, 0.09 and 0.06 for 200, 600, and 1000 weeks.

in the estimation increases. This is quite different from the case of the constant covariance structure, where more data is always helpful in improving the estimation of a covariance matrix. This shows that when the covariance structure is time-varying, there is a trade-off between the sampling variation and the data being obsolete.[14] Therefore, there will be an optimal amount of data for estimation.

We therefore plot Figure 7.1 to show that relationship of the RMSE and eigen-distance results with the amount of the time series data used in the covariance estimation. We can see that many estimators exhibit a slight U-shape performance, the bottom of the curve indicates the optimal amount of data for the estimation. We can see that the optimal number of observations for different estimators vary between T=200 to 250 weeks.

We can see from Figure 7.1 that for RiskMetrics, the curve under the eigen-distance measure decreases then stays almost flat. This is the consequence of its exponential weighting scheme that regardless of the amount of observations, the effective number of data used in its estimation is limited by the size of the decaying factor.[15] We also consider a RiskMetrics estimator with a different decaying factor of $\lambda = 0.99$ (RM(0.99)). It performs much better and rather differently. We also find that the sample historical covariance matrix (H) improves most with increasing data under both measures.

---

[14]Except for methods like RiskMetrics that weights the observations.

[15]For example, if we use λ=0.95, then 99.9% information is contained in the last 135 ($=log(0.001)/log(0.95)$) observations.

7.4 Conclusion

This chapter extends our analysis in Chapter 6 to explore improvements on covariance estimation under two more realistic situations: a covariance structure for returns with fat-tailed distributions and a time-varying covariance structure where stocks migrate among different risk classes. We find that our eigenvalue shrinkage estimators are still very useful in both circumstances, and become even more important in the time varying case. On the other hand, we find that methods imposing more weights on the recent information and taking an average cross sectional stock relationships, which are expected to be useful in the time varying case, did not do better.

The time-varying covariance structure raises many issues that need to be explored further. Our work in this chapter has explored how things change in the time varying situation, which is important to understand the properties of these time variations.

Table 7.1: **RMSE and RMSE differences (MNIG T=200 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 200 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 6.1 in Chapter 6. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
H	1.9854	0.1630	H	2.9023	0.1869
S	1.8643	0.1685	S	2.8394	0.1899
V	1.8274	0.1613	TV	2.7920	0.1908
SV	1.8207	0.1615	JM	2.7816	0.1907
SVlog	1.8189	0.1616	LW	2.7786	0.1899
LW	1.8177	0.1648	V	2.7744	0.1885
TV	1.8100	0.1685	SV	2.7698	0.1886
JM	1.7828	0.1692	SVlog	2.7690	0.1886
Panel B: RMSE Differences					
	mean	se t-stats		mean	se t-stats
S - H	-0.1211	0.0155 (-7.81 *)	S - H	-0.0629	0.0133 (-4.72 *)
LW - H	-0.1677	0.0106 (-15.82 *)	LW - H	-0.1237	0.0098 (-12.60 *)
LW - S	-0.0466	0.0143 (-3.27 *)	LW - S	-0.0609	0.0159 (-3.83 *)
JM - H	-0.2026	0.0135 (-15.00 *)	JM - H	-0.1207	0.0096 (-12.61 *)
JM - S	-0.0815	0.0058 (-14.09 *)	JM - S	-0.0578	0.0058 (-10.03 *)
JM - LW	-0.0349	0.0105 (-3.33 *)	JM - LW	0.0030	0.0112 (0.27)
V - H	-0.1580	0.0173 (-9.11 *)	V - H	-0.1280	0.0163 (-7.86 *)
V - S	-0.0369	0.0199 (-1.86)	V - S	-0.0651	0.0224 (-2.90 *)
V - LW	0.0097	0.0145 (0.67)	V - LW	-0.0042	0.0107 (-0.39)
V - JM	0.0446	0.0183 (2.44 *)	V - JM	-0.0072	0.0184 (-0.39)
SV - H	-0.1647	0.0179 (-9.20 *)	SV - H	-0.1325	0.0168 (-7.87 *)
SV - S	-0.0436	0.0198 (-2.20 *)	SV - S	-0.0696	0.0225 (-3.09 *)
SV - LW	0.0030	0.0148 (0.20)	SV - LW	-0.0087	0.0110 (-0.80)
SV - JM	0.0379	0.0183 (2.07 *)	SV - JM	-0.0117	0.0186 (-0.63)
SV - V	-0.0067	0.0014 (-4.81 *)	SV - V	-0.0045	0.0016 (-2.78 *)
SVlog- H	-0.1665	0.0180 (-9.24 *)	SVlog- H	-0.1333	0.0168 (-7.92 *)
SVlog- S	-0.0454	0.0197 (-2.30 *)	SVlog- S	-0.0704	0.0224 (-3.15 *)
SVlog- LW	0.0012	0.0149 (0.08)	SVlog- LW	-0.0096	0.0110 (-0.88)
SVlog- JM	0.0361	0.0183 (1.98 *)	SVlog- JM	-0.0126	0.0185 (-0.68)
SVlog- V	-0.0085	0.0018 (-4.73 *)	SVlog- V	-0.0054	0.0019 (-2.88 *)
SVlog- SV	-0.0018	0.0005 (-3.68 *)	SVlog- SV	-0.0009	0.0003 (-2.64 *)
TV - H	-0.1754	0.0128 (-13.69 *)	TV - H	-0.1103	0.0094 (-11.71 *)
TV - S	-0.0543	0.0116 (-4.67 *)	TV - S	-0.0474	0.0110 (-4.33 *)
TV - LW	-0.0077	0.0101 (-0.76)	TV - LW	0.0134	0.0091 (1.48)
TV - JM	0.0272	0.0089 (3.07 *)	TV - JM	0.0104	0.0071 (1.47)
TV - V	-0.0173	0.0146 (-1.19)	TV - V	0.0177	0.0159 (1.11)
TV - SV	-0.0107	0.0147 (-0.73)	TV - SV	0.0222	0.0160 (1.38)
TV - SVlog	-0.0089	0.0147 (-0.61)	TV - SVlog	0.0230	0.0159 (1.44)

Table 7.2: **Eigen-distance and eigen-distance differences (MNIG T=200 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators. when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 200 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
H	3.4028	0.0260	H	4.8450	0.0241
S	2.3645	0.0135	S	4.1322	0.0287
LW	2.2923	0.0279	LW	4.0906	0.0288
V	2.1272	0.0222	V	3.9605	0.0273
SVlog	2.1184	0.0223	TV	3.9485	0.0265
SV	2.1181	0.0218	SV	3.9469	0.0277
TV	2.0975	0.0216	SVlog	3.9461	0.0276
JM	2.0040	0.0228	JM	3.9008	0.0261

Panel B: Eigen-distance Differences

	mean	se	t-stats		mean	se	t-stats
S - H	-1.0383	0.0269	(-38.66 *)	S - H	-0.7128	0.0301	(-23.65 *)
LW - H	-1.1105	0.0197	(-56.45 *)	LW - H	-0.7543	0.0165	(-45.77 *)
LW - S	-0.0722	0.0259	(-2.79 *)	LW - S	-0.0415	0.0302	(-1.38)
JM - H	-1.3988	0.0189	(-74.02 *)	JM - H	-0.9442	0.0176	(-53.70 *)
JM - S	-0.3605	0.0211	(-17.06 *)	JM - S	-0.2314	0.0212	(-10.91 *)
JM - LW	-0.2884	0.0179	(-16.15 *)	JM - LW	-0.1899	0.0154	(-12.29 *)
V - H	-1.2756	0.0296	(-43.05 *)	V - H	-0.8845	0.0299	(-29.54 *)
V - S	-0.2373	0.0186	(-12.78 *)	V - S	-0.1717	0.0259	(-6.62 *)
V - LW	-0.1651	0.0264	(-6.25 *)	V - LW	-0.1301	0.0279	(-4.66 *)
V - JM	0.1232	0.0263	(4.69 *)	V - JM	0.0597	0.0226	(2.64 *)
SV - H	-1.2847	0.0310	(-41.49 *)	SV - H	-0.8981	0.0312	(-28.74 *)
SV - S	-0.2464	0.0194	(-12.72 *)	SV - S	-0.1853	0.0266	(-6.97 *)
SV - LW	-0.1742	0.0265	(-6.58 *)	SV - LW	-0.1437	0.0291	(-4.93 *)
SV - JM	0.1141	0.0271	(4.22 *)	SV - JM	0.0461	0.0237	(1.95)
SV - V	-0.0091	0.0074	(-1.22)	SV - V	-0.0136	0.0059	(-2.30 *)
SVlog- H	-1.2844	0.0313	(-40.99 *)	SVlog- H	-0.8989	0.0314	(-28.63 *)
SVlog- S	-0.2462	0.0198	(-12.41 *)	SVlog- S	-0.1861	0.0265	(-7.03 *)
SVlog- LW	-0.1740	0.0265	(-6.57 *)	SVlog- LW	-0.1445	0.0292	(-4.95 *)
SVlog- JM	0.1144	0.0274	(4.18 *)	SVlog- JM	0.0453	0.0238	(1.91)
SVlog- V	-0.0088	0.0086	(-1.03)	SVlog- V	-0.0144	0.0058	(-2.46 *)
SVlog- SV	0.0002	0.0026	(0.10)	SVlog- SV	-0.0008	0.0014	(-0.54)
TV - H	-1.3053	0.0302	(-43.23 *)	TV - H	-0.8964	0.0296	(-30.34 *)
TV - S	-0.2670	0.0178	(-14.98 *)	TV - S	-0.1837	0.0243	(-7.54 *)
TV - LW	-0.1948	0.0299	(-6.53 *)	TV - LW	-0.1421	0.0282	(-5.04 *)
TV - JM	0.0935	0.0248	(3.77 *)	TV - JM	0.0477	0.0215	(2.22 *)
TV - V	-0.0297	0.0116	(-2.56 *)	TV - V	-0.0120	0.0063	(-1.91)
TV - SV	-0.0206	0.0152	(-1.36)	TV - SV	0.0016	0.0087	(0.19)
TV - SVlog	-0.0209	0.0162	(-1.29)	TV - SVlog	0.0024	0.0086	(0.28)

Table 7.3: **RMSE and RMSE differences (MNIG T=600 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 600 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
S	1.3313	0.0714	S	1.7361	0.0833
H	1.1923	0.0721	H	1.6410	0.0850
LW	1.1536	0.0720	LW	1.6131	0.0855
SV	1.1439	0.0721	SV	1.6126	0.0852
V	1.1431	0.0721	SVlog	1.6120	0.0853
TV	1.1428	0.0741	V	1.6113	0.0853
SVlog	1.1427	0.0722	TV	1.6076	0.0862
JM	1.1392	0.0752	JM	1.6006	0.0867
Panel B: RMSE Differences					
	mean	se t-stats		mean	se t-stats
S - H	0.1390	0.0096 (14.45 *)	S - H	0.0950	0.0117 (8.15 *)
LW - H	-0.0387	0.0037 (-10.42 *)	LW - H	-0.0279	0.0037 (-7.62 *)
LW - S	-0.1777	0.0115 (-15.48 *)	LW - S	-0.1229	0.0133 (-9.26 *)
JM - H	-0.0531	0.0077 (-6.86 *)	JM - H	-0.0405	0.0072 (-5.61 *)
JM - S	-0.1921	0.0070 (-27.51 *)	JM - S	-0.1355	0.0069 (-19.60 *)
JM - LW	-0.0144	0.0083 (-1.73)	JM - LW	-0.0126	0.0082 (-1.54)
V - H	-0.0493	0.0058 (-8.43 *)	V - H	-0.0297	0.0064 (-4.65 *)
V - S	-0.1882	0.0107 (-17.64 *)	V - S	-0.1247	0.0134 (-9.27 *)
V - LW	-0.0105	0.0059 (-1.78)	V - LW	-0.0018	0.0062 (-0.29)
V - JM	0.0038	0.0085 (0.45)	V - JM	0.0108	0.0093 (1.17)
SV - H	-0.0485	0.0060 (-8.03 *)	SV - H	-0.0285	0.0065 (-4.39 *)
SV - S	-0.1874	0.0108 (-17.43 *)	SV - S	-0.1235	0.0134 (-9.23 *)
SV - LW	-0.0097	0.0060 (-1.62)	SV - LW	-0.0006	0.0062 (-0.09)
SV - JM	0.0046	0.0085 (0.54)	SV - JM	0.0120	0.0092 (1.30)
SV - V	0.0008	0.0008 (0.96)	SV - V	0.0012	0.0008 (1.49)
SVlog- H	-0.0496	0.0061 (-8.18 *)	SVlog- H	-0.0291	0.0065 (-4.47 *)
SVlog- S	-0.1886	0.0107 (-17.56 *)	SVlog- S	-0.1241	0.0133 (-9.30 *)
SVlog- LW	-0.0109	0.0061 (-1.80)	SVlog- LW	-0.0012	0.0063 (-0.19)
SVlog- JM	0.0035	0.0085 (0.41)	SVlog- JM	0.0114	0.0092 (1.25)
SVlog- V	-0.0004	0.0011 (-0.35)	SVlog- V	0.0006	0.0009 (0.70)
SVlog- SV	-0.0012	0.0003 (-3.84 *)	SVlog- SV	-0.0006	0.0002 (-3.24 *)
TV - H	-0.0496	0.0041 (-12.06 *)	TV - H	-0.0335	0.0036 (-9.20 *)
TV - S	-0.1886	0.0097 (-19.52 *)	TV - S	-0.1285	0.0111 (-11.60 *)
TV - LW	-0.0109	0.0049 (-2.24 *)	TV - LW	-0.0056	0.0047 (-1.19)
TV - JM	0.0035	0.0063 (0.56)	TV - JM	0.0070	0.0061 (1.15)
TV - V	-0.0003	0.0052 (-0.06)	TV - V	-0.0038	0.0054 (-0.70)
TV - SV	-0.0011	0.0053 (-0.21)	TV - SV	-0.0050	0.0054 (-0.93)
TV - SVlog	0.0000	0.0052 (0.01)	TV - SVlog	-0.0044	0.0053 (-0.83)

Table 7.4: **Eigen-distance and eigen-distance differences (MNIG T=600 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 600 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
S	2.2786	0.0040	S	2.8697	0.0177
H	1.8406	0.0147	H	2.5768	0.0158
LW	1.5725	0.0139	LW	2.3830	0.0162
SVlog	1.4710	0.0133	SV	2.3041	0.0161
SV	1.4709	0.0134	SVlog	2.3028	0.0162
TV	1.4483	0.0117	V	2.2932	0.0155
V	1.4460	0.0124	TV	2.2928	0.0155
JM	1.4144	0.0128	JM	2.2356	0.0144
Panel B: Eigen-distance Differences					
	mean	se t-stats		mean	se t-stats
S - H	0.4380	0.0146 (30.10 *)	S - H	0.2929	0.0211 (13.85 *)
LW - H	-0.2681	0.0073 (-36.75 *)	LW - H	-0.1938	0.0067 (-29.07 *)
LW - S	-0.7062	0.0143 (-49.52 *)	LW - S	-0.4867	0.0209 (-23.28 *)
JM - H	-0.4263	0.0119 (-35.92 *)	JM - H	-0.3412	0.0127 (-26.84 *)
JM - S	-0.8643	0.0131 (-66.16 *)	JM - S	-0.6342	0.0152 (-41.76 *)
JM - LW	-0.1581	0.0098 (-16.18 *)	JM - LW	-0.1474	0.0112 (-13.16 *)
V - H	-0.3946	0.0165 (-23.95 *)	V - H	-0.2836	0.0154 (-18.38 *)
V - S	-0.8326	0.0124 (-67.35 *)	V - S	-0.5765	0.0176 (-32.77 *)
V - LW	-0.1265	0.0148 (-8.57 *)	V - LW	-0.0897	0.0148 (-6.05 *)
V - JM	0.0316	0.0143 (2.22 *)	V - JM	0.0577	0.0118 (4.88 *)
SV - H	-0.3697	0.0180 (-20.53 *)	SV - H	-0.2727	0.0167 (-16.32 *)
SV - S	-0.8078	0.0137 (-59.11 *)	SV - S	-0.5656	0.0175 (-32.32 *)
SV - LW	-0.1016	0.0159 (-6.39 *)	SV - LW	-0.0788	0.0158 (-4.98 *)
SV - JM	0.0565	0.0156 (3.62 *)	SV - JM	0.0686	0.0127 (5.41 *)
SV - V	0.0249	0.0067 (3.70 *)	SV - V	0.0109	0.0053 (2.06 *)
SVlog- H	-0.3697	0.0182 (-20.32 *)	SVlog- H	-0.2740	0.0168 (-16.32 *)
SVlog- S	-0.8077	0.0135 (-59.70 *)	SVlog- S	-0.5669	0.0175 (-32.39 *)
SVlog- LW	-0.1016	0.0160 (-6.33 *)	SVlog- LW	-0.0802	0.0160 (-5.02 *)
SVlog- JM	0.0566	0.0157 (3.60 *)	SVlog- JM	0.0672	0.0129 (5.22 *)
SVlog- V	0.0249	0.0068 (3.69 *)	SVlog- V	0.0096	0.0052 (1.84)
SVlog- SV	0.0001	0.0017 (0.03)	SVlog- SV	-0.0013	0.0008 (-1.59)
TV - H	-0.3924	0.0154 (-25.51 *)	TV - H	-0.2840	0.0153 (-18.60 *)
TV - S	-0.8304	0.0115 (-72.52 *)	TV - S	-0.5769	0.0177 (-32.56 *)
TV - LW	-0.1242	0.0143 (-8.68 *)	TV - LW	-0.0902	0.0148 (-6.10 *)
TV - JM	0.0339	0.0132 (2.57 *)	TV - JM	0.0572	0.0116 (4.93 *)
TV - V	0.0023	0.0036 (0.62)	TV - V	-0.0005	0.0019 (-0.24)
TV - SV	-0.0226	0.0072 (-3.15 *)	TV - SV	-0.0113	0.0059 (-1.93)
TV - SVlog	-0.0227	0.0074 (-3.08 *)	TV - SVlog	-0.0100	0.0058 (-1.72)

Table 7.5: **RMSE and RMSE differences (MNIG T=1000 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 1000 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: RMSE					
	mean	se		mean	se
S	1.0940	0.0538	S	1.4100	0.0600
H	0.8582	0.0559	JM	1.2225	0.0642
JM	0.8549	0.0569	H	1.2161	0.0641
LW	0.8415	0.0560	LW	1.2033	0.0647
TV	0.8347	0.0566	TV	1.1986	0.0648
SV	0.8344	0.0561	SV	1.1949	0.0650
SVlog	0.8330	0.0561	V	1.1943	0.0650
V	0.8320	0.0561	SVlog	1.1939	0.0650

Panel B: RMSE Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.2358	0.0092	(25.64 *)	S - H	0.1939	0.0108	(18.00 *)
LW - H	-0.0167	0.0024	(-7.01 *)	LW - H	-0.0128	0.0025	(-5.18 *)
LW - S	-0.2525	0.0107	(-23.59 *)	LW - S	-0.2066	0.0124	(-16.69 *)
JM - H	-0.0033	0.0063	(-0.53)	JM - H	0.0064	0.0057	(1.14)
JM - S	-0.2391	0.0063	(-38.20 *)	JM - S	-0.1874	0.0070	(-26.81 *)
JM - LW	0.0134	0.0071	(1.88)	JM - LW	0.0192	0.0069	(2.78 *)
V - H	-0.0263	0.0035	(-7.43 *)	V - H	-0.0218	0.0037	(-5.88 *)
V - S	-0.2621	0.0100	(-26.23 *)	V - S	-0.2156	0.0122	(-17.61 *)
V - LW	-0.0095	0.0038	(-2.48 *)	V - LW	-0.0090	0.0035	(-2.58 *)
V - JM	-0.0229	0.0067	(-3.41 *)	V - JM	-0.0282	0.0068	(-4.12 *)
SV - H	-0.0238	0.0038	(-6.25 *)	SV - H	-0.0212	0.0040	(-5.33 *)
SV - S	-0.2596	0.0100	(-25.99 *)	SV - S	-0.2151	0.0122	(-17.66 *)
SV - LW	-0.0071	0.0040	(-1.79)	SV - LW	-0.0084	0.0037	(-2.28 *)
SV - JM	-0.0205	0.0067	(-3.07 *)	SV - JM	-0.0276	0.0068	(-4.07 *)
SV - V	0.0024	0.0008	(2.89 *)	SV - V	0.0006	0.0009	(0.60)
SVlog- H	-0.0253	0.0039	(-6.50 *)	SVlog- H	-0.0222	0.0040	(-5.51 *)
SVlog- S	-0.2611	0.0100	(-26.14 *)	SVlog- S	-0.2161	0.0122	(-17.75 *)
SVlog- LW	-0.0085	0.0040	(-2.11 *)	SVlog- LW	-0.0094	0.0038	(-2.51 *)
SVlog- JM	-0.0219	0.0066	(-3.31 *)	SVlog- JM	-0.0286	0.0068	(-4.22 *)
SVlog- V	0.0010	0.0010	(1.03)	SVlog- V	-0.0005	0.0011	(-0.43)
SVlog- SV	-0.0014	0.0002	(-6.56 *)	SVlog- SV	-0.0010	0.0002	(-5.59 *)
TV - H	-0.0235	0.0026	(-9.13 *)	TV - H	-0.0176	0.0022	(-7.92 *)
TV - S	-0.2593	0.0094	(-27.64 *)	TV - S	-0.2114	0.0106	(-19.91 *)
TV - LW	-0.0067	0.0034	(-2.00 *)	TV - LW	-0.0048	0.0032	(-1.47)
TV - JM	-0.0202	0.0059	(-3.40 *)	TV - JM	-0.0240	0.0051	(-4.70 *)
TV - V	0.0028	0.0031	(0.90)	TV - V	0.0042	0.0033	(1.28)
TV - SV	0.0003	0.0032	(0.11)	TV - SV	0.0036	0.0034	(1.08)
TV - SVlog	0.0018	0.0032	(0.55)	TV - SVlog	0.0047	0.0034	(1.37)

Table 7.6: **Eigen-distance and eigen-distance differences (MNIG T=1000 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eight alternative covariance estimators, when compared against both the true covariance matrix and an ex-post sample of the true covariance matrix. The length of time-series observation is 1000 weeks. Results in the two sections of Panel A are ranked separately according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. * denotes statistical significance at a 95% confidence level.

vs. true			out-of-sample		
Panel A: Eigen-distance					
	mean	se		mean	se
S	2.2653	0.0028	S	2.6861	0.0158
H	1.4163	0.0127	H	1.9946	0.0134
LW	1.2796	0.0120	LW	1.9061	0.0132
JM	1.2702	0.0105	JM	1.8792	0.0137
SVlog	1.2062	0.0101	SV	1.8484	0.0144
SV	1.2060	0.0100	SVlog	1.8467	0.0143
TV	1.1810	0.0096	V	1.8436	0.0137
V	1.1787	0.0098	TV	1.8435	0.0138

Panel B: Eigen-distance Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.8490	0.0125	(68.00 *)	S - H	0.6915	0.0196	(35.27 *)
LW - H	-0.1367	0.0041	(-33.23 *)	LW - H	-0.0885	0.0035	(-25.58 *)
LW - S	-0.9857	0.0121	(-81.38 *)	LW - S	-0.7800	0.0196	(-39.86 *)
JM - H	-0.1461	0.0111	(-13.20 *)	JM - H	-0.1154	0.0142	(-8.13 *)
JM - S	-0.9951	0.0104	(-95.80 *)	JM - S	-0.8069	0.0145	(-55.61 *)
JM - LW	-0.0094	0.0091	(-1.03)	JM - LW	-0.0269	0.0124	(-2.17 *)
V - H	-0.2376	0.0139	(-17.07 *)	V - H	-0.1510	0.0165	(-9.17 *)
V - S	-1.0866	0.0092	(-118.01 *)	V - S	-0.8425	0.0186	(-45.33 *)
V - LW	-0.1009	0.0132	(-7.63 *)	V - LW	-0.0626	0.0153	(-4.09 *)
V - JM	-0.0915	0.0118	(-7.77 *)	V - JM	-0.0357	0.0134	(-2.67 *)
SV - H	-0.2103	0.0149	(-14.15 *)	SV - H	-0.1462	0.0177	(-8.26 *)
SV - S	-1.0593	0.0098	(-107.65 *)	SV - S	-0.8377	0.0192	(-43.54 *)
SV - LW	-0.0737	0.0142	(-5.19 *)	SV - LW	-0.0577	0.0166	(-3.48 *)
SV - JM	-0.0642	0.0129	(-4.98 *)	SV - JM	-0.0308	0.0145	(-2.12 *)
SV - V	0.0273	0.0054	(5.07 *)	SV - V	0.0049	0.0047	(1.04)
SVlog- H	-0.2101	0.0149	(-14.14 *)	SVlog- H	-0.1479	0.0178	(-8.31 *)
SVlog- S	-1.0592	0.0100	(-106.18 *)	SVlog- S	-0.8394	0.0192	(-43.64 *)
SVlog- LW	-0.0735	0.0141	(-5.20 *)	SVlog- LW	-0.0595	0.0167	(-3.57 *)
SVlog- JM	-0.0641	0.0129	(-4.97 *)	SVlog- JM	-0.0326	0.0146	(-2.24 *)
SVlog- V	0.0274	0.0056	(4.88 *)	SVlog- V	0.0031	0.0046	(0.67)
SVlog- SV	0.0002	0.0013	(0.15)	SVlog- SV	-0.0018	0.0009	(-2.04 *)
TV - H	-0.2353	0.0138	(-17.05 *)	TV - H	-0.1511	0.0164	(-9.20 *)
TV - S	-1.0843	0.0090	(-120.83 *)	TV - S	-0.8426	0.0184	(-45.76 *)
TV - LW	-0.0986	0.0132	(-7.48 *)	TV - LW	-0.0627	0.0153	(-4.10 *)
TV - JM	-0.0892	0.0118	(-7.58 *)	TV - JM	-0.0357	0.0134	(-2.67 *)
TV - V	0.0023	0.0016	(1.44)	TV - V	-0.0001	0.0014	(-0.06)
TV - SV	-0.0250	0.0057	(-4.35 *)	TV - SV	-0.0050	0.0051	(-0.98)
TV - SVlog	-0.0252	0.0059	(-4.27 *)	TV - SVlog	-0.0032	0.0050	(-0.63)

Table 7.7: **RMSE and RMSE differences (OU, α=0.03, T=200 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=201 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The two additional short forms are: RM for RiskMetrics estimator and C for average covariance estimator. * denotes statistical significance at a 95% confidence level.

Panel A: RMSE		
	mean	se
RiskMetrics	2.4615	0.0302
C	1.6072	0.0069
RM+C	1.5172	0.0138
RM+S	1.4924	0.0172
S	1.2525	0.0098
H	1.2277	0.0084
LW	1.1159	0.0084
JM	1.1017	0.0099
V	1.0929	0.0094
SVlog	1.0912	0.0095
TV	0.9946	0.0094

Panel B: RMSE Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.0248	0.0040	(6.27 *)	RM - H	1.2338	0.0314	(39.27 *)
LW - H	-0.1118	0.0032	(-34.46 *)	RM - S	1.2090	0.0314	(38.55 *)
LW - S	-0.1366	0.0057	(-24.11 *)	RM - LW	1.3457	0.0317	(42.41 *)
JM - H	-0.1260	0.0028	(-44.90 *)	RM - JM	1.3599	0.0317	(42.87 *)
JM - S	-0.1508	0.0020	(-75.89 *)	RM - V	1.3687	0.0324	(42.24 *)
JM - LW	-0.0142	0.0045	(-3.15 *)	RM - SVLOG	1.3703	0.0324	(42.29 *)
V - H	-0.1348	0.0112	(-12.07 *)	RM - TV	1.4669	0.0315	(46.58 *)
V - S	-0.1596	0.0130	(-12.27 *)	RM - C	0.8544	0.0311	(27.45 *)
V - LW	-0.0230	0.0106	(-2.17 *)	RMS - H	0.2646	0.0178	(14.90 *)
V - JM	-0.0088	0.0125	(-0.71)	RMS - S	0.2398	0.0177	(13.51 *)
SVlog- H	-0.1365	0.0112	(-12.19 *)	RMS - LW	0.3765	0.0180	(20.90 *)
SVlog- S	-0.1613	0.0130	(-12.38 *)	RMS - JM	0.3907	0.0180	(21.68 *)
SVlog- LW	-0.0247	0.0106	(-2.33 *)	RMS - V	0.3995	0.0200	(19.98 *)
SVlog- JM	-0.0105	0.0125	(-0.84)	RMS - SVlog	0.4012	0.0200	(20.06 *)
SVlog- V	-0.0017	0.0002	(-7.61 *)	RMS - TV	0.4977	0.0196	(25.38 *)
TV - H	-0.2331	0.0125	(-18.65 *)	RMS - C	-0.1148	0.0179	(-6.42 *)
TV - S	-0.2579	0.0134	(-19.25 *)	RMS - RM	-0.9692	0.0154	(-62.73 *)
TV - LW	-0.1213	0.0123	(-9.85 *)	RMC - H	0.2895	0.0148	(19.56 *)
TV - JM	-0.1071	0.0133	(-8.03 *)	RMC - S	0.2647	0.0157	(16.90 *)
TV - V	-0.0983	0.0071	(-13.92 *)	RMC - LW	0.4013	0.0141	(28.47 *)
TV - SVlog	-0.0966	0.0071	(-13.63 *)	RMC - JM	0.4155	0.0156	(26.65 *)
C - H	0.3795	0.0081	(46.59 *)	RMC - V	0.4243	0.0153	(27.66 *)
C - S	0.3547	0.0095	(37.51 *)	RMC - SVLOG	0.4260	0.0154	(27.73 *)
C - LW	0.4913	0.0066	(73.91 *)	RMC - TV	0.5226	0.0157	(33.34 *)
C - JM	0.5055	0.0093	(54.13 *)	RMC - C	-0.0900	0.0134	(-6.72 *)
C - V	0.5143	0.0109	(47.38 *)	RMC - RM	-0.9444	0.0217	(-43.57 *)
C - SVLOG	0.5160	0.0109	(47.37 *)	RMC - RMS	0.0248	0.0093	(2.67 *)
C - TV	0.6126	0.0117	(52.25 *)				

Table 7.8: **Eigen-distance and eigen-distance differences (OU, α=0.03, T=200 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=201 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 7.7. * denotes statistical significance at a 95% confidence level.

Panel A: Eigen-distance

	mean	se
RiskMetrics	6.7021	0.0152
H	3.0397	0.0106
C	2.7441	0.0090
RM+C	2.7253	0.0094
RM+S	2.7233	0.0090
S	2.5805	0.0098
LW	2.2952	0.0091
JM	2.0641	0.0088
SVlog	1.9961	0.0077
V	1.9930	0.0076
TV	1.9524	0.0076

Panel B: Eigen-distance Differences

	mean	se	t-stats		mean	se	t-stats
S - H	-0.4593	0.0123	(-37.35 *)	RM - H	3.6623	0.0178	(206.22 *)
LW - H	-0.7446	0.0093	(-79.78 *)	RM - S	4.1216	0.0175	(236.02 *)
LW - S	-0.2853	0.0096	(-29.77 *)	RM - LW	4.4069	0.0163	(270.89 *)
JM - H	-0.9756	0.0110	(-88.31 *)	RM - JM	4.6380	0.0165	(281.64 *)
JM - S	-0.5164	0.0067	(-77.30 *)	RM - V	4.7091	0.0172	(273.83 *)
JM - LW	-0.2311	0.0071	(-32.57 *)	RM - SVLOG	4.7060	0.0172	(274.09 *)
V - H	-1.0468	0.0119	(-87.73 *)	RM - TV	4.7497	0.0164	(290.01 *)
V - S	-0.5875	0.0119	(-49.51 *)	RM - C	3.9579	0.0166	(238.59 *)
V - LW	-0.3022	0.0099	(-30.38 *)	RMS - H	-0.3165	0.0143	(-22.08 *)
V - JM	-0.0711	0.0088	(-8.07 *)	RMS - S	0.1428	0.0128	(11.16 *)
SVlog- H	-1.0437	0.0120	(-87.15 *)	RMS - LW	0.4281	0.0122	(35.01 *)
SVlog- S	-0.5844	0.0120	(-48.74 *)	RMS - JM	0.6592	0.0115	(57.53 *)
SVlog- LW	-0.2991	0.0100	(-29.90 *)	RMS - V	0.7303	0.0116	(62.94 *)
SVlog- JM	-0.0680	0.0090	(-7.56 *)	RMS - SVlog	0.7272	0.0118	(61.41 *)
SVlog- V	0.0031	0.0021	(1.46)	RMS - TV	0.7709	0.0111	(69.62 *)
TV - H	-1.0874	0.0119	(-91.16 *)	RMS - C	-0.0208	0.0124	(-1.68)
TV - S	-0.6281	0.0114	(-54.89 *)	RMS - RM	-3.9788	0.0147	(-270.94 *)
TV - LW	-0.3428	0.0098	(-34.87 *)	RMC - H	-0.3145	0.0146	(-21.60 *)
TV - JM	-0.1117	0.0090	(-12.42 *)	RMC - S	0.1448	0.0131	(11.04 *)
TV - V	-0.0406	0.0053	(-7.71 *)	RMC - LW	0.4301	0.0125	(34.53 *)
TV - SVlog	-0.0437	0.0056	(-7.81 *)	RMC - JM	0.6612	0.0117	(56.40 *)
C - H	-0.2956	0.0122	(-24.27 *)	RMC - V	0.7323	0.0118	(62.19 *)
C - S	0.1637	0.0091	(17.97 *)	RMC - SVLOG	0.7292	0.0120	(60.52 *)
C - LW	0.4489	0.0092	(48.90 *)	RMC - TV	0.7729	0.0114	(67.84 *)
C - JM	0.6800	0.0092	(73.93 *)	RMC - C	-0.0188	0.0124	(-1.52)
C - V	0.7511	0.0111	(67.91 *)	RMC - RM	-3.9768	0.0142	(-280.82 *)
C - SVLOG	0.7480	0.0110	(68.06 *)	RMC - RMS	0.0020	0.0038	(0.53)
C - TV	0.7917	0.0105	(75.67 *)				

Table 7.9: **RMSE and RMSE differences (OU, α=0.03, T=600 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=601 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 7.7. * denotes statistical significance at a 95% confidence level.

Panel A: RMSE		
	mean	se
RiskMetrics	2.4819	0.0305
C	1.6716	0.0070
RM+C	1.5194	0.0127
RM+S	1.4539	0.0156
S	1.3096	0.0051
V	1.1351	0.0086
SVlog	1.1336	0.0087
JM	1.1278	0.0054
H	1.1273	0.0052
LW	1.1057	0.0051
TV	1.0503	0.0092

Panel B: RMSE Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.1823	0.0027	(68.39 *)	RM - H	1.3545	0.0305	(44.42 *)
LW - H	-0.0216	0.0011	(-19.10 *)	RM - S	1.1722	0.0303	(38.68 *)
LW - S	-0.2039	0.0030	(-68.41 *)	RM - LW	1.3762	0.0305	(45.05 *)
JM - H	0.0005	0.0015	(0.33)	RM - JM	1.3540	0.0304	(44.58 *)
JM - S	-0.1818	0.0015	(-122.72 *)	RM - V	1.3467	0.0323	(41.64 *)
JM - LW	0.0221	0.0020	(11.19 *)	RM - SVLOG	1.3483	0.0323	(41.71 *)
V - H	0.0078	0.0082	(0.94)	RM - TV	1.4316	0.0329	(43.55 *)
V - S	-0.1745	0.0084	(-20.88 *)	RM - C	0.8102	0.0313	(25.91 *)
V - LW	0.0294	0.0083	(3.54 *)	RMS - H	0.3266	0.0154	(21.27 *)
V - JM	0.0073	0.0083	(0.88)	RMS - S	0.1443	0.0152	(9.52 *)
SVlog- H	0.0063	0.0083	(0.76)	RMS - LW	0.3482	0.0154	(22.63 *)
SVlog- S	-0.1760	0.0084	(-20.96 *)	RMS - JM	0.3261	0.0152	(21.46 *)
SVlog- LW	0.0279	0.0083	(3.34 *)	RMS - V	0.3188	0.0181	(17.64 *)
SVlog- JM	0.0058	0.0083	(0.69)	RMS - SVlog	0.3203	0.0181	(17.73 *)
SVlog- V	-0.0015	0.0002	(-8.82 *)	RMS - TV	0.4036	0.0189	(21.33 *)
TV - H	-0.0771	0.0112	(-6.87 *)	RMS - C	-0.2177	0.0167	(-13.02 *)
TV - S	-0.2593	0.0109	(-23.88 *)	RMS - RM	-1.0280	0.0164	(-62.75 *)
TV - LW	-0.0554	0.0111	(-5.00 *)	RMC - H	0.3920	0.0129	(30.35 *)
TV - JM	-0.0775	0.0112	(-6.91 *)	RMC - S	0.2097	0.0129	(16.23 *)
TV - V	-0.0848	0.0072	(-11.76 *)	RMC - LW	0.4137	0.0126	(32.77 *)
TV - SVlog	-0.0833	0.0072	(-11.55 *)	RMC - JM	0.3915	0.0130	(30.17 *)
C - H	0.5443	0.0077	(70.81 *)	RMC - V	0.3842	0.0161	(23.82 *)
C - S	0.3620	0.0075	(48.17 *)	RMC - SVLOG	0.3858	0.0161	(23.92 *)
C - LW	0.5659	0.0069	(82.59 *)	RMC - TV	0.4691	0.0166	(28.32 *)
C - JM	0.5438	0.0077	(70.99 *)	RMC - C	-0.1523	0.0126	(-12.05 *)
C - V	0.5365	0.0111	(48.46 *)	RMC - RM	-0.9625	0.0220	(-43.74 *)
C - SVLOG	0.5380	0.0111	(48.49 *)	RMC - RMS	0.0655	0.0088	(7.43 *)
C - TV	0.6213	0.0113	(55.18 *)				

Table 7.10: **Eigen-distance and eigen-distance differences (OU, α=0.03, T=600 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=601 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 7.7. * denotes statistical significance at a 95% confidence level.

Panel A: Eigen-distance

	mean	se
RiskMetrics	6.7236	0.0154
C	3.0414	0.0090
S	2.9541	0.0098
RM+C	2.7779	0.0090
RM+S	2.7610	0.0089
H	2.5150	0.0115
LW	2.4320	0.0119
JM	2.4283	0.0119
SVlog	2.2330	0.0106
V	2.2318	0.0103
TV	2.2140	0.0109

Panel B: Eigen-distance Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.4391	0.0112	(39.37 *)	RM - H	4.2086	0.0191	(220.69 *)
LW - H	-0.0831	0.0019	(-44.48 *)	RM - S	3.7695	0.0180	(209.87 *)
LW - S	-0.5222	0.0110	(-47.26 *)	RM - LW	4.2917	0.0193	(221.90 *)
JM - H	-0.0868	0.0065	(-13.36 *)	RM - JM	4.2954	0.0194	(221.56 *)
JM - S	-0.5259	0.0077	(-68.35 *)	RM - V	4.4919	0.0181	(247.60 *)
JM - LW	-0.0037	0.0056	(-0.66)	RM - SVLOG	4.4906	0.0184	(243.45 *)
V - H	-0.2833	0.0047	(-60.58 *)	RM - TV	4.5096	0.0183	(246.32 *)
V - S	-0.7224	0.0107	(-67.24 *)	RM - C	3.6822	0.0172	(214.56 *)
V - LW	-0.2002	0.0047	(-42.58 *)	RMS - H	0.2460	0.0132	(18.64 *)
V - JM	-0.1965	0.0067	(-29.48 *)	RMS - S	-0.1931	0.0120	(-16.10 *)
SVlog- H	-0.2820	0.0043	(-66.31 *)	RMS - LW	0.3290	0.0135	(24.35 *)
SVlog- S	-0.7211	0.0107	(-67.11 *)	RMS - JM	0.3327	0.0134	(24.89 *)
SVlog- LW	-0.1989	0.0042	(-46.84 *)	RMS - V	0.5292	0.0128	(41.20 *)
SVlog- JM	-0.1952	0.0064	(-30.27 *)	RMS - SVlog	0.5279	0.0130	(40.58 *)
SVlog- V	0.0013	0.0015	(0.83)	RMS - TV	0.5470	0.0133	(41.12 *)
TV - H	-0.3010	0.0053	(-57.10 *)	RMS - C	-0.2804	0.0115	(-24.49 *)
TV - S	-0.7401	0.0114	(-64.72 *)	RMS - RM	-3.9626	0.0153	(-259.47 *)
TV - LW	-0.2179	0.0050	(-43.40 *)	RMC - H	0.2629	0.0137	(19.12 *)
TV - JM	-0.2142	0.0071	(-30.35 *)	RMC - S	-0.1762	0.0126	(-14.00 *)
TV - V	-0.0177	0.0037	(-4.82 *)	RMC - LW	0.3459	0.0140	(24.67 *)
TV - SVlog	-0.0190	0.0040	(-4.73 *)	RMC - JM	0.3496	0.0139	(25.08 *)
C - H	0.5264	0.0131	(40.30 *)	RMC - V	0.5461	0.0133	(41.21 *)
C - S	0.0873	0.0097	(9.02 *)	RMC - SVLOG	0.5449	0.0134	(40.65 *)
C - LW	0.6095	0.0129	(47.40 *)	RMC - TV	0.5639	0.0136	(41.40 *)
C - JM	0.6131	0.0121	(50.50 *)	RMC - C	-0.2635	0.0108	(-24.38 *)
C - V	0.8097	0.0125	(64.87 *)	RMC - RM	-3.9457	0.0148	(-266.88 *)
C - SVLOG	0.8084	0.0126	(64.19 *)	RMC - RMS	0.0169	0.0037	(4.63 *)
C - TV	0.8274	0.0129	(64.30 *)				

Table 7.11: **RMSE and RMSE differences (OU, α=0.03, T=1000 weeks)** This table reports the 100 simulations results of the RMSE (Panel A) and RMSE differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=1001 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 7.7. * denotes statistical significance at a 95% confidence level.

Panel A: RMSE		
	mean	se
RiskMetrics	2.4900	0.0349
C	1.6927	0.0084
RM+C	1.5301	0.0140
RM+S	1.4598	0.0166
S	1.4041	0.0061
V	1.2497	0.0065
JM	1.2484	0.0064
SVlog	1.2480	0.0065
H	1.2396	0.0062
LW	1.2251	0.0061
TV	1.0890	0.0103

Panel B: RMSE Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.1645	0.0023	(70.37 *)	RM - H	1.2504	0.0353	(35.46 *)
LW - H	-0.0145	0.0006	(-23.10 *)	RM - S	1.0858	0.0353	(30.77 *)
LW - S	-0.1790	0.0024	(-73.87 *)	RM - LW	1.2648	0.0354	(35.75 *)
JM - H	0.0088	0.0013	(6.56 *)	RM - JM	1.2416	0.0353	(35.15 *)
JM - S	-0.1558	0.0013	(-121.51 *)	RM - V	1.2403	0.0357	(34.78 *)
JM - LW	0.0232	0.0015	(16.02 *)	RM - SVLOG	1.2419	0.0357	(34.82 *)
V - H	0.0101	0.0053	(1.89)	RM - TV	1.4009	0.0372	(37.71 *)
V - S	-0.1545	0.0057	(-26.92 *)	RM - C	0.7973	0.0373	(21.39 *)
V - LW	0.0245	0.0052	(4.68 *)	RMS - H	0.2202	0.0169	(13.04 *)
V - JM	0.0013	0.0056	(0.23)	RMS - S	0.0557	0.0170	(3.28 *)
SVlog- H	0.0084	0.0053	(1.59)	RMS - LW	0.2347	0.0170	(13.82 *)
SVlog- S	-0.1561	0.0057	(-27.34 *)	RMS - JM	0.2115	0.0170	(12.46 *)
SVlog- LW	0.0229	0.0052	(4.39 *)	RMS - V	0.2102	0.0176	(11.97 *)
SVlog- JM	-0.0003	0.0056	(-0.06)	RMS - SVlog	0.2118	0.0176	(12.06 *)
SVlog- V	-0.0016	0.0001	(-12.10 *)	RMS - TV	0.3708	0.0200	(18.56 *)
TV - H	-0.1506	0.0119	(-12.70 *)	RMS - C	-0.2329	0.0194	(-12.03 *)
TV - S	-0.3151	0.0117	(-27.01 *)	RMS - RM	-1.0301	0.0196	(-52.51 *)
TV - LW	-0.1361	0.0117	(-11.67 *)	RMC - H	0.2905	0.0142	(20.51 *)
TV - JM	-0.1594	0.0119	(-13.41 *)	RMC - S	0.1260	0.0144	(8.72 *)
TV - V	-0.1607	0.0094	(-17.10 *)	RMC - LW	0.3050	0.0142	(21.52 *)
TV - SVlog	-0.1590	0.0094	(-16.86 *)	RMC - JM	0.2818	0.0144	(19.63 *)
C - H	0.4531	0.0088	(51.71 *)	RMC - V	0.2805	0.0149	(18.77 *)
C - S	0.2886	0.0084	(34.38 *)	RMC - SVLOG	0.2821	0.0149	(18.87 *)
C - LW	0.4676	0.0082	(56.71 *)	RMC - TV	0.4411	0.0176	(25.06 *)
C - JM	0.4443	0.0086	(51.68 *)	RMC - C	-0.1625	0.0162	(-10.03 *)
C - V	0.4430	0.0092	(47.91 *)	RMC - RM	-0.9598	0.0246	(-38.95 *)
C - SVLOG	0.4446	0.0092	(48.08 *)	RMC - RMS	0.0703	0.0085	(8.27 *)
C - TV	0.6037	0.0109	(55.61 *)				

Table 7.12: **Eigen-distance and eigen-distance differences (OU, α=0.03, T=1000 weeks)** This table reports the 100 simulations results of the eigen-distance (Panel A) and eigen-distance differences (Panel B) of eleven alternative covariance estimators, when compared against the *instant true* covariance matrix at T=1001 weeks. Results in Panel A are ranked according to the ascending orders of the performance of models. Results in Panel B are reported in the fixed order. The short forms are the same as explained in Table 7.7. * denotes statistical significance at a 95% confidence level.

Panel A: Eigen-distance		
	mean	se
RiskMetrics	6.7334	0.0163
C	3.1459	0.0100
S	3.0685	0.0094
RM+C	2.8064	0.0090
RM+S	2.7921	0.0090
H	2.6368	0.0129
LW	2.5985	0.0129
JM	2.5912	0.0117
SVlog	2.4491	0.0115
V	2.4445	0.0113
TV	2.4218	0.0119

Panel B: Eigen-distance Differences

	mean	se	t-stats		mean	se	t-stats
S - H	0.4317	0.0106	(40.67 *)	RM - H	4.0966	0.0206	(198.86 *)
LW - H	-0.0383	0.0010	(-38.81 *)	RM - S	3.6649	0.0191	(191.95 *)
LW - S	-0.4700	0.0105	(-44.93 *)	RM - LW	4.1349	0.0206	(201.08 *)
JM - H	-0.0457	0.0054	(-8.42 *)	RM - JM	4.1423	0.0199	(208.58 *)
JM - S	-0.4774	0.0075	(-63.27 *)	RM - V	4.2890	0.0203	(211.25 *)
JM - LW	-0.0073	0.0049	(-1.50)	RM - SVLOG	4.2843	0.0204	(210.31 *)
V - H	-0.1924	0.0055	(-34.89 *)	RM - TV	4.3116	0.0206	(209.27 *)
V - S	-0.6241	0.0105	(-59.52 *)	RM - C	3.5875	0.0186	(193.33 *)
V - LW	-0.1541	0.0054	(-28.29 *)	RMS - H	0.1553	0.0148	(10.52 *)
V - JM	-0.1467	0.0067	(-22.03 *)	RMS - S	-0.2764	0.0132	(-20.99 *)
SVlog- H	-0.1878	0.0051	(-36.61 *)	RMS - LW	0.1936	0.0147	(13.17 *)
SVlog- S	-0.6195	0.0105	(-58.82 *)	RMS - JM	0.2010	0.0138	(14.60 *)
SVlog- LW	-0.1494	0.0050	(-29.67 *)	RMS - V	0.3477	0.0133	(26.16 *)
SVlog- JM	-0.1421	0.0064	(-22.15 *)	RMS - SVlog	0.3430	0.0135	(25.42 *)
SVlog- V	0.0046	0.0017	(2.74 *)	RMS - TV	0.3703	0.0137	(27.09 *)
TV - H	-0.2150	0.0058	(-36.80 *)	RMS - C	-0.3537	0.0135	(-26.20 *)
TV - S	-0.6467	0.0102	(-63.50 *)	RMS - RM	-3.9413	0.0163	(-241.59 *)
TV - LW	-0.1767	0.0056	(-31.37 *)	RMC - H	0.1696	0.0148	(11.45 *)
TV - JM	-0.1694	0.0066	(-25.61 *)	RMC - S	-0.2621	0.0137	(-19.12 *)
TV - V	-0.0227	0.0039	(-5.81 *)	RMC - LW	0.2079	0.0148	(14.09 *)
TV - SVlog	-0.0273	0.0041	(-6.68 *)	RMC - JM	0.2153	0.0140	(15.33 *)
C - H	0.5090	0.0134	(38.05 *)	RMC - V	0.3620	0.0134	(27.08 *)
C - S	0.0773	0.0097	(7.99 *)	RMC - SVLOG	0.3573	0.0136	(26.34 *)
C - LW	0.5474	0.0131	(41.78 *)	RMC - TV	0.3846	0.0139	(27.75 *)
C - JM	0.5547	0.0116	(47.64 *)	RMC - C	-0.3394	0.0133	(-25.49 *)
C - V	0.7014	0.0122	(57.47 *)	RMC - RM	-3.9270	0.0163	(-240.75 *)
C - SVLOG	0.6968	0.0124	(56.12 *)	RMC - RMS	0.0143	0.0035	(4.13 *)
C - TV	0.7241	0.0123	(58.86 *)				

Chapter 8

Conclusion

This thesis has studied several issues related to the estimation of a covariance matrix for the returns of a reasonably large number of stocks for portfolio risk management. Chapter 3 provides a comprehensive examination on the covariance estimation methods and the standard comparison criteria. Chapter 4 and 5 introduce robust alternative appraisal methods of covariance estimators for portfolio risk management purposes. Chapter 6 and 7 explore improvements on the best covariance estimators within the literature. The key results and findings are summarized in this chapter. We also point out areas for potential future research.

8.1 Summary and conclusion

Chapter 3 provides a comprehensive analysis of both old and new covariance estimation methods and the standard comparison criteria used in the related research. Using

empirical data, we examine the relative performance of these estimation methods under the two conventional comparison criteria, namely, the RMSE and the volatility analysis of the minimum variance portfolio. We find that other than a few common results, the two comparison criteria in general give systematically different results. In short, methods that impose richer structures or adjusting for noises tend to do better under the MVP test than under the RMSE measure, and the opposite is true for methods that impose simpler structures. As the RMSE is a statistical measure and the MVP test provides only limited information regarding the ability of covariance estimators in predicting portfolio variances based on one special portfolio, we need to explore more powerful comparison criteria in order to better assess different covariance estimators for portfolio risk management purposes. This empirical analysis also shows that we need to use simulations to better understanding the sampling properties of difference covariance estimators.

Motivated by the need to search for a robust alternative comparison criterion, Chapter 4 proposes a portfolio distance measure based on eigen-decomposition (*eigen-distance*) to compare alternative estimates of a covariance matrix in terms of the biggest differences of portfolio variances they predict. We show that this innovative *eigen-distance* measure has sound interpretations on both statistical and economic grounds and therefore can be used as a single measure to compare different covariance estimators. This helps to eliminate the problems of inconsistency when two measures

are used separately.

More importantly, we show that the *eigen-distance* is suitable for evaluating a risk system as a whole, where any kind of the portfolios may need to be considered. It can be applied equally well to both the absolute variances of portfolios and the variances of their tracking errors against a benchmark. Our simulation results show it is a powerful measure to distinguish two covariance estimators even in small samples. This is very useful as in practice the amount of time series observations available for forecasting is usually limited.

Chapter 5 proposes a θ measure to distinguish two similar estimated covariance matrices from the observed covariance matrix. θ is based on the essential difference of the two estimated covariance matrices, i.e., the performance of the two extreme portfolios that are predicted to have the most different variances under these two covariance matrices. It is designed to measure how much two covariance matrices differ and whether one matrix is a clear improvement of the other. As a modest refinement usually results in the refined covariance matrix being relatively close to the original covariance matrix, our θ measure is extremely useful in refining covariance estimators.

Our simulation results show that the θ measure is more powerful than both the RMSE and eigen-distance measures in differentiating two similar covariance matrices. In addition, the relative power of θ depends on the sampling size as well as the relative

closeness of the two matrices to the observed covariance matrix. For two reasonably close covariance matrices, θ is very powerful in small samples. When both matrices are a long way from the true covariance structure, however, the θ results (not surprisingly) may not give helpful signals.

Chapter 6 is concerned with improving the best covariance estimators within the literature. A number of recent studies have found that the Ledoit and Wolf (2003a) estimator and the Jagannathan and Ma (2003) estimator perform better than many other covariance estimators. We explore alternative Bayesian shrinkage estimators based on directly shrinking the eigenvalues of the sample covariance matrix (and in one case shrinking the principal eigenvector as well), and compare their performance with these two covariance estimators using both the RMSE and eigen-distance criteria. We find that we have succeed to a considerable extent. Our simulation results show that our shrinkage estimators consistently beat the Ledoit and Wolf estimator. They also out-perform the Jagannathan and Ma estimator by a considerable amount most of the time except in one case, where they are not much worse than the Jagannathan and Ma estimator.

Finally, Chapter 7 extends the analysis of Chapter 6 of an unchanging multivariate normal world to explore the implications of both fat tails and time variation of stock returns. In the first extension, we use a multivariate normal inverse Gaussian (MNIG) distribution to model the log returns of stock prices. This family of distributions has

proven to fit the heavy tails observed in financial time series extremely well. In the second extension, we consider a more interesting and economically motivated time varying covariance structure where the general market risk characteristics are constant while stocks migrate among different risk categories during their life cycles. We develop an original model that employs a tractable mean reverting Ornstein-Uhlenbeck (OU) process to model the time varying factor loadings in a multivariate factor model. This gives us a situation where the cross-section of risk characteristics remains unchanged but the identities of stocks do. Our model is very useful to explore and understand how things change in the time varying situation. The simulation results show that our shrinkage methods are still very useful in both circumstances and become even more important in the time varying case.

8.2 Limitations

The eigen-distance measure we have introduced can be applied directly to the correlation matrices or normalized covariance matrices. We assumed no covariance matrices are a scalar multiple of any other covariance matrices, and used this eigen-distance measure without any normalization on the covariance matrices in this thesis. This greatly simplified the derivations and enabled us to establish a more powerful economic interpretation for the measure. We define the eigen-distance

as $d(\mathbf{V}_1, \mathbf{V}_2) = ln\left(\frac{\lambda_{max}}{\lambda_{min}}\right)$. If the scaling issue was thought to be a major problem, we could instead define for example $d(\mathbf{V}_1, \mathbf{V}_2) = |ln(\lambda_{max})| + |ln(\lambda_{min})|$, or $d(\mathbf{V}_1, \mathbf{V}_2) = Max\{0, +ln(\lambda_{max})\} + Max\{0, -ln(\lambda_{min})\}$, which satisfies Condition 2 ($d(x, y) = 0$ if and only if $x = y$), but does not satisfy Condition 3 (triangle inequality).[1] These measures are different only when either $\lambda_{min} > 1$ or $\lambda_{max} < 1$, thus virtually never happens in our numerical work. In situations where λ_{min} and λ_{max} are more likely to lie on the same side of 1, the appropriate choice of metric would be less clear cut.

Much of our simulation analysis is based on 78 liquid NYSE stocks over a particular period of time, which could potentially have influenced our results and limited the interpretation of these results to a particular covariance structure. However, we expect most observed properties to carry over but with the modifications noted below. The choice of the time period is less particularly significant than the number of stocks, although the covariance structure may also be different at other time periods. Our covariance matrix is characterized in terms of various underlying factors (market factor, industry factor and residuals). By choosing large stocks, we make the factors (in particular the market factor) stronger, and idiosyncratic risks less strong. We can estimate such a stronger structure using less amount of data. For smaller and less liquid stocks, we expect the estimation to be less precise and affect the estimators

[1]The conditions we present for a distance function are the standard axioms for a metric space (Copson (1968)).

less. Further empirical studies however are necessary in order for us to draw any conclusions regarding the relative performance of different covariance estimators for different sets of stocks, time periods and markets.

This thesis studies the estimation problem of a covariance matrix. We use simulations to compare how different the estimated covariance matrices are from a true covariance matrix. In Chapter 4, Chapter 6 and first extension of Chapter 7 where we assume a world with a constant true covariance structure, we report the comparison of the estimated covariance matrices with an ex-post sample of the true covariance matrix to show that it is more difficult to estimate a covariance matrix when there is sampling variation. In the second extension (time variation) of Chapter 7, since it is much harder to measure the sampling variation, we concentrated on reporting the comparisons of estimated covariance matrices with the true covariance matrix. In Chapter 5, we design a measure to differentiate two similar covariance estimators.

8.3 Future research

There are a number of areas that can be extended in the future research. First, in Chapter 4, we have derived a Corollary for the weights of the two extreme portfolios that are predicted to have the most different variances under two different covariance matrices (Corollary 4.2.1). It will be interesting to explore the characteristics of these extreme portfolios for different types of the covariance estimators. For example, we

may use these portfolios to construct variance swaps for people to bet on different risk models.

Second, we have studied the covariance matrix of returns for a fixed 78 stocks with a few pre-specified lengths of time series observations. Our results clearly depend on both the real covariance structure and the ratio of the number of stocks to the amount of time series observations in our sample, N/T. It is worth investigating in more detail how N/T affects our results.

Third, we have studied the case of no weight constraints as we are interested in the maximum forecasting differences of two covariance estimators and this gives us more tractable results when we introduce the two new alternative measures. A potential area for further study is to consider some portfolio constraints to see if we can improve our results by concentrating on a sub-set of stocks in the portfolio.

Chapter 7 raises many issues that need to be explored further. In particular, in a time-varying world, it seems natural to put more weight on recent observations than on older ones and to use covariance estimators based on average relationships. We find these models do not do better in our numerical experiment. This shows that we need to do more empirical work to calibrate our time varying model. In addition, as we have seen, we need to further explore adjustments which are vital to preserve an appropriate eigenvalue structure in this kind of situations. We have only scratched the surface of how to construct better covariance estimators in such situations.

Finally, our study focuses on the covariance estimation methods that use mostly historical return data. Many practitioners also use methods, for example, the BARRA risk systems, to incorporate more fundamental data. We may include these types of methods in future research.

Appendix A

Company information

In our empirical studies, we use the Datastream's Wednesday to Wednesday weekly dividend-adjusted stock prices for the NYSE US 100 index constituent companies as of 11 May 2006 from 06 January 1988 to 25 December 1996. In total, 78 stocks have the full historical data we require. The table below summarizes the names and the NYSE economic sector classification of these stocks. The NYSE sectors have four levels of classifications: economic sector, market sector, industry group and sub-group. In this thesis, we only use the first level classification (the economic sector), to consider broad industry groups effect.

Table A.1: **Stock names and their industry classifications**

Name	Sector	Name	Sector
Alcoa	Basic Materials	Eli Lilly	Health Care
EI DuPont de Nemours	Basic Materials	Merck	Health Care
Dow Chemical	Basic Materials	Pfizer	Health Care
Anheuser-Busch Cos	Consumer Goods	Schering-Plough	Health Care
Coca-Cola	Consumer Goods	Wyeth	Health Care
PepsiCo	Consumer Goods	Boeing	Industrials
Procter Gamble	Consumer Goods	United Technologies	Industrials
Colgate-Palmolive	Consumer Goods	Emerson Electric	Industrials
Kimberly-Clark	Consumer Goods	General Electric	Industrials
Altria Group	Consumer Goods	Honeywell Intl	Industrials
Walt Disney	Consumer Services	3M	Industrials
Cardinal Health	Consumer Services	Caterpillar	Industrials
Walgreen	Consumer Services	Illinois Tool Works	Industrials
Home Depot	Consumer Services	Burlington Northern Santa Fe	Industrials
Lowe's Cos	Consumer Services	Automatic Data Processing	Industrials
Target	Consumer Services	First Data	Industrials
Wal-Mart Stores	Consumer Services	ConocoPhillips	Oil Gas
Carnival	Consumer Services	Chevron	Oil Gas
McDonald's	Consumer Services	Occidental Petroleum	Oil Gas
Bank of America	Financials	Valero Energy	Oil Gas
Bank of New York	Financials	Exxon Mobil	Oil Gas
Citigroup	Financials	Baker Hughes	Oil Gas
JPMorgan Chase	Financials	Halliburton	Oil Gas
National City	Financials	Schlumberger	Oil Gas
SunTrust Banks	Financials	Intl Business Machines	Technology
US Bank	Financials	EMC	Technology
Wachovia	Financials	Corning	Technology
Wells Fargo	Financials	Hewlett-Packard	Technology
Washington Mutual	Financials	Motorola	Technology
American Express	Financials	Texas Instruments	Technology
Fannie Mae	Financials	BellSouth	Telecommunications
Merrill Lynch	Financials	ATT	Telecommunications
American Intl Group	Financials	Verizon Communications	Telecommunications
St Paul Travelers Cos	Financials	Sprint Nextel	Telecommunications
Medtronic	Health Care	Dominion Resources (Virginia)	Utilities
UnitedHealth Group	Health Care	Exelon	Utilities
Abbott Laboratories	Health Care	Southern	Utilities
Bristol-Myers Squibb	Health Care	TXU	Utilities
Johnson Johnson	Health Care	Duke Energy	Utilities

Appendix B

Additional summary statistics of variables in Chapter 3

This section reports the summary statistics of the variables used in the covariance estimators compared in Chapter 3.

(1) We calculate the cross-sectional average estimated betas for the first and second sub-sample period of 01.1988-12.1990 and 01.1991-12.1993, the realized beta for the ex-post period of 01.1994-12.1996, the Blume (1971) adjusted beta forecast, and the Vasicek (1973) adjusted beta forecast.

	1st sub-period	2nd sub-period	realized ex-post	Blume adjusted	Vasicek adjusted
Mean	0.8717	0.9884	0.8984	1.1015	0.9832
Std	0.2391	0.3731	0.3096	0.3620	0.0193

We can see that the Blume (1971) adjustment dose not work well for our sample

data. In particular, because on average the estimated betas in the second sub-period is higher than that of the first sub-period, the Blume adjustment picks up the upward trend from the first and second sub-periods and extrapolates the trend. However, the realized beta in the ex-post period is lower than that of the second sub-period, indicating probably a mean-reverting pattern rather than a continued upward trend.

We also consider a more general Blume type adjustment, which assumes that $\hat{\beta}_j = \bar{\beta} + k \times (\beta_{2,j} - \bar{\beta}_{2,j})$, where $\bar{\beta} = 0.5 \times (\bar{\beta}_1 + \bar{\beta}_2)$. $k = \rho \times \frac{\sigma_{\beta_3}}{\sigma_{\beta_2}}$, where we assume $\sigma_{\beta_3} = \sqrt{0.5 \times (\sigma^2_{\beta_1} + \sigma^2_{\beta_2})}$ and ρ is the correlation coefficient of $\beta_{1,j}$ and $\beta_{2,j}$. We find that in this case the average adjusted beta equals 0.9300 and its standard deviation equals 0.1948, which are better than those obtained using the Blume (1971) adjustment.

(2) The Blume (1971) linear regression has an intercept equals 0.1427 and coefficient equals 0.9701.

(3) We use a simple graphical method called the scree test proposed by Cattell (1966) to justify the choice of five principal components in our multi-index model.

The scree test involves plotting the eigenvalues in a simple line plot and finding the place where the smooth decrease of eigenvalues appears to level off to the right of the plot. We can see from Figure (B.1) that in our example, the eigenvalues drop sharply starting with the sixth largest eigenvalues.

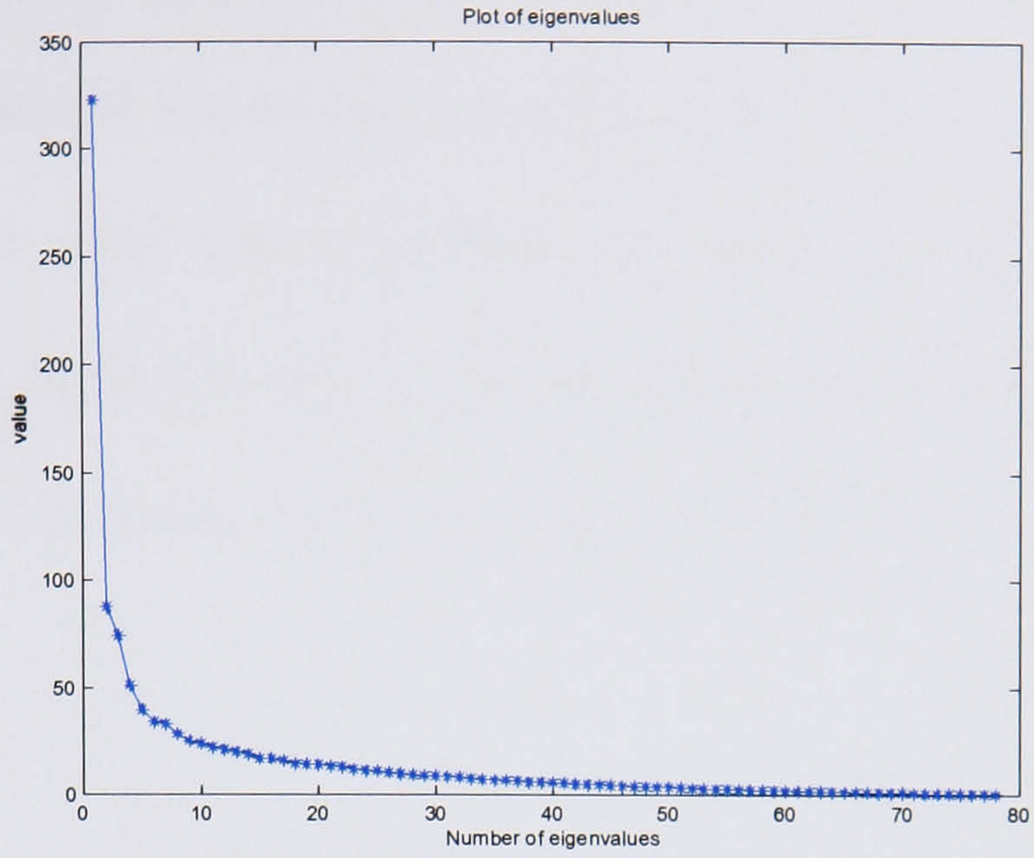


Figure B.1: **Scree test to determine the number of principal components factors** This figure plots in descending orders the eigenvalues of our sample historical covariance matrix of 78 stocks for the period of 01.1991 to 12.1993 (see Section 3.4 for details of our sample data). Five principal components are selected using Cattell's criterion.

(4) As the RiskMetrics (1996) uses 0.94 and 0.97 for the daily and monthly estimation, a reasonable guess is that the optimal decay factor for weekly data will be somewhere between 0.94 and 0.97. We arbitrarily choose $\lambda = 0.95$ for our study, although we could have calculated the exact value of the optimal decaying factor for our data sample.

(5) For the Ledoit and Wolf (2003a) estimator, the shrinkage intensity or the weight placed on the single-index covariance estimator equals 0.2509.

(6) For the random matrix filtering method, the upper bound of the eigenvalues

predicted by the random matrix theory for the correlation matrix of our second sub-period with 157 weeks of observations and 78 stocks equals 2.9065. There are five eigenvalues that are above this threshold value and they equal: 21.0010, 5.7506, 4.4219, 3.4727 and 2.3249. The remaining eigenvalues are all averaged to equal 0.5859 in order to maintain the trace of the correlation matrix.

Appendix C

Proof of $d(\mathbf{V_1}, \mathbf{V_2})$ as a distance function in Chapter 4

We prove in the following that our portfolio-based eigen-distance $d(\mathbf{V_1}, \mathbf{V_2})$ satisfies all four required conditions as a proper distance function. These conditions are:

1. $d(x, y) \geq 0$ (*non-negativity*)

2. $d(x, y) = 0$ if and only if x is equivalent y (*identity of indiscernible*)

   (where we regard $x = y$ if and only if $x = ky$ for some k)

3. $d(x, y) = d(y, x)$ (*symmetry*)

4. $d(x, z) \leq d(x, y) + d(y, z)$ (*triangle inequality*)

Non-negativity

Since $d(\mathbf{V_1}, \mathbf{V_2})$ measures the logarithm value, which is always positive, the non-negativity condition is satisfied.

Identity of indiscernible

$\mathbf{V_1}$ is equivalent to $\mathbf{V_2}$ if and only if $\mathbf{V_1} = k\mathbf{V_2}$ for some k, in which case, $d(\mathbf{V_1}, \mathbf{V_2}) = \log(\frac{k}{k}) = 0$. $d(\mathbf{V_1}, \mathbf{V_2})$ can be applied directly to the correlation matrices or normalized matrices (for example, where they have the same sum of all matrix elements). We apply this distance measure even without such normalization.

Symmetry

Let $\max_{\mathbf{x}} \frac{\mathbf{x'V_1x}}{\mathbf{x'V_2x}} = a$ and $\min_{\mathbf{y}} \frac{\mathbf{y'V_1y}}{\mathbf{y'V_2y}} = b$, then

$$\max_{\mathbf{x}} \frac{\mathbf{x'V_2x}}{\mathbf{x'V_1x}} = 1/b$$

$$\min_{\mathbf{y}} \frac{\mathbf{y'V_2y}}{\mathbf{y'V_1y}} = 1/a$$

$$d(\mathbf{V_1}, \mathbf{V_2}) = \log\left(\frac{a}{b}\right)$$

$$d(\mathbf{V_2}, \mathbf{V_1}) = \log\left(\frac{1/b}{1/a}\right) = \log\left(\frac{a}{b}\right) = d(\mathbf{V_1}, \mathbf{V_2})$$

Triangle inequality

Suppose there are three covariance matrices $\mathbf{V_1}$, $\mathbf{V_2}$ and $\mathbf{V_3}$, the triangle inequality

requires that we proof $d(\mathbf{V_1}, \mathbf{V_2}) + d(\mathbf{V_2}, \mathbf{V_3}) \geq d(\mathbf{V_1}, \mathbf{V_3})$.

$$
\begin{aligned}
d(\mathbf{V_1}, \mathbf{V_2}) + d(\mathbf{V_2}, \mathbf{V_3}) &= \log\left(\frac{\max_{\mathbf{x_{12}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}}}{\min_{\mathbf{y_{12}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}}}\right) + \log\left(\frac{\max_{\mathbf{x_{23}}} \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{23}}} \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \\
&= \log\left(\frac{\max_{\mathbf{x_{12}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}}}{\min_{\mathbf{y_{12}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}}} \times \frac{\max_{\mathbf{x_{23}}} \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{23}}} \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \\
&= \log\left(\frac{\max_{\mathbf{x_{12},x_{23}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}} \times \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{12},y_{23}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}} \times \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \qquad \text{(C.0.1)} \\
d(\mathbf{V_1}, \mathbf{V_3}) &= \log\left(\frac{\max_{\mathbf{x_{13}}} \frac{\mathbf{x'_{13}V_1x_{13}}}{\mathbf{x'_{13}V_3x_{13}}}}{\min_{\mathbf{y_{13}}} \frac{\mathbf{y'_{13}V_1y_{13}}}{\mathbf{y'_{13}V_3y_{13}}}}\right) \qquad \text{(C.0.2)} \\
&= \log\left(\frac{\max_{\mathbf{x_{12},x_{23}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}} \times \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{12},y_{23}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}} \times \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \text{ s.t.} \begin{matrix}\mathbf{x_{12}=x_{23}} \\ \mathbf{y_{12}=y_{23}}\end{matrix}
\end{aligned}
$$

The above calculation shows that $d(\mathbf{V_1}, \mathbf{V_2}) + d(\mathbf{V_2}, \mathbf{V_3})$ is the log ratio of the unconstrained maximization and minimization, while $d(\mathbf{V_1}, \mathbf{V_3})$ is the log ratio of the constrained maximization and minimization. Since the result of an unconstrained maximization must be greater than or equal to that of a constrained maximization, Equation (C.0.1) is always greater than Equation (C.0.2), i.e.,

$$
\log\left(\frac{\max_{\mathbf{x_{12},x_{23}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}} \times \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{12},y_{23}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}} \times \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \geq \log\left(\frac{\max_{\mathbf{x_{12},x_{23}}} \frac{\mathbf{x'_{12}V_1x_{12}}}{\mathbf{x'_{12}V_2x_{12}}} \times \frac{\mathbf{x'_{23}V_2x_{23}}}{\mathbf{x'_{23}V_3x_{23}}}}{\min_{\mathbf{y_{12},y_{23}}} \frac{\mathbf{y'_{12}V_1y_{12}}}{\mathbf{y'_{12}V_2y_{12}}} \times \frac{\mathbf{y'_{23}V_2y_{23}}}{\mathbf{y'_{23}V_3y_{23}}}}\right) \text{ s.t.} \begin{matrix}\mathbf{x_{12}=x_{23}} \\ \mathbf{y_{12}=y_{23}}\end{matrix} \qquad \text{(C.0.3)}
$$

therefore

$$
d(\mathbf{V_1}, \mathbf{V_2}) + d(\mathbf{V_2}, \mathbf{V_3}) \geq d(\mathbf{V_1}, \mathbf{V_3})
$$

Bibliography

[1] O.E. Barndorff-Nielsen. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24:1–14, 1997a.

[2] O.E. Barndorff-Nielsen. Processes of normal inverse gaussian type. *Finance and Stochastics*, 2:41–68, 1997b.

[3] O.E. Barndorff-Nielsen and Shephard N. Non-gaussian ornstein-uhlenbeck based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, 63:167–241, 2001.

[4] G.K. Basak, R. Jagannathan, and T. Ma. Assessing the risk in sample minimum risk portfolios. NBER working paper No. W10447, 2004.

[5] C. Bengtsson and J. Holst. On portfolio selection: improved covariance matrix estimation for swedish asset returns. Working paper, Department of Economics, Lund University, 2002.

[6] F.E. Benth, M. Groth, and P. Kettler. A quasi-monte carlo algorithm for the normal inverse gaussian distribution and valuation of financial derivatives. *International Journal of Theoretical and Applied Finance*, 9:843–867, 2006.

[7] M.J. Best and R.R. Grauer. Sensitivity analysis for mean-variance portfolio problems. *Management Science*, 37:980–989, 1991.

[8] M.E. Blume. On the assessment of risk. *Journal of Finance*, 26:1–10, 1971.

[9] M.E. Blume. Beta and their regression tendencies. *Journal of Finance*, 30:785–795, 1975.

[10] T. Bollerslev, R.F. Engle, and J.M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

[11] S. Brown. The number of factors in security returns. *Journal of Finance*, 44:1247–1262, 1989.

[12] J.Y. Campbell. Understanding risk and return. *Journal of Political Economy*, 104:298–345, 1996.

[13] J.Y. Campbell, M. Lettau, B.G. Malkiel, and Y. Xu. Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *Journal of Finance*, 56:1–43, 2001.

[14] R.B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1:245–276, 1966.

[15] L.K.C. Chan, J. Karceski, and J. Lakonishok. On portfolio optimization: forecasting covariance and choosing the risk model. *Review of Financial Studies*. 12:937–974, 1999.

[16] N. Chen, R. Ross, and S. Roll. Economic forces and the stock market. *Journal of Business*, 59:383–403, 1986.

[17] V.K. Chopra and W.T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.

[18] T. Chordia and B. Swaminathan. Trading volumn and cross-autocorrelations in stock returns. *Journal of Finance*, 55:913–935, 2000.

[19] K.J. Cohen and J.A. Pogue. An empirical evaluation of alternative portfolio-section models. *Journal of Business*, 40:166–193, 1967.

[20] G. Connor. The three types of factor models: a comparison of their explanatory power. *Financial Analyst Journal*, 51:42–46, 1995.

[21] G. Connor and R.A. Korajczyk. A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–1291, 1993.

[22] E.T. Copson. *Metric spaces.* Cambridge University Press, New York, 1968.

[23] M.J. Daniels and R.E. Kass. Shrinkage estimators for covariance matrices. *Biometrics*, 57:1173–1184, 2001.

[24] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies, forthcoming*, 2005.

[25] D.K. Dey and C. Srinivasan. Estimation of a covariance matrix under stein's loss. *Annals of Statistics*, 13:1581–1591, 1985.

[26] D.J. Disatnik and S. Benninga. Estimating the covariance matrix for portfolio optimization. Working paper, Tel Aviv University, Israel, 2005.

[27] E. Eberlein and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1:281–299, 1995.

[28] E.J. Elton and M.J. Gruber. Estimating the dependence structure of share prices - implications for portfolio selection. *Journal of Finance*, 28:1203–1232, 1973.

[29] E.J. Elton, M.J. Gruber, and J. Spitzer. Improved estimates of correlation coefficients and their impact on the optimum portfolios. *European Financial Management*, 12, 2006.

[30] E.J. Elton, M.J. Gruber, and T.J. Urich. Are betas best? *Journal of Finance*, 33:1375–1384, 1978.

[31] E.J. Elton, M.J. Martin, S.J. Brown, and W.N. Goetzmann. *Modern portfolio theory and investment analysis.* Wiley, 7th edition, 2007.

[32] R.F. Engle and K.F. Kroner. Multivariate simultaneous generolized arch. *Econometric Theory*, 11:122–150, 1995.

[33] C.S. Eun and B.G. Resnick. Estimating the correlation structure of international share prices. *Journal of Finance*, 39:1311–1324, 1984.

[34] C.S. Eun and B.G. Resnick. Forecasting the correlation structure of share prices: a test of new models. *Journal of Banking and Finance*, 16:643–656, 1992.

[35] E.F. Fama and K.R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

[36] W. Ferson and C. Harvey. The variation of economic risk premium. *Journal of Political Economy*, 99:385–415, 1991.

[37] M. Giliberto. Interest rate sensitivity in the common stocks of financial interme-diaries: a methodological note. *Journal of Financial and Quantitative Analysis.* 20:123–126, 1985.

[38] L.R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics*, 8:586–597, 1980.

[39] L.R. Haff. The variational form of certain bayes estimators. *Annals of Statistics*, 19:1163–1190, 1991.

[40] J.E. Jackson. *A user's guide to principal component.* Wiley - Interscience, 1991.

[41] R. Jagannathan and T. Ma. Risk reduction in large portfolios: why imposing the wrong constrains helps. *Journal of Finance*, 58:1651–1683, 2003.

[42] R. Jagannathan and Z. Wang. The conditional capm and the cross-section of expected returns. *Journal of Finance*, 51:3–53, 1996.

[43] W. James and C. Stein. Estimation with quadratic loss. In L.M. LeCam and J. Neyman, editors, *Proceedings of the Fourth Berkeley Symposium on Mathematical and Statistical Probablity*, volume 1, pages 361–379. University of California, Berkeley, 1961.

[44] N.L. Johnson and S. Kotz. *Distributions in statistics: Continuous multivariate distributions.* Wiley, New York, London, 1972.

[45] B. King. Market and industry factors in stock price behavior. *Journal of Business*, 39:139–190, 1966.

[46] R. Klemkosky and J. Martin. The adjustment of beta forecasts. *Journal of Finance*, 10:1123–1128, 1975.

[47] K.F. Kroner and V.K. Ng. Modeling asymmetric comovements of asset returns. *Review of Financial Studies*, 11:817–844, 1998.

[48] L. Laloux, P. Cizeau, and M. Potters. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3:391–397, 2000.

[49] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud. Random matrix theory and financial correlations. *Physical Review Letter*, 83:1469, 1999.

[50] O. Ledoit, P. Santa-Clara, and M. Wolf. Flexible multivariate garch modelling with an application to international stock markets. *Review of Economics and Statistics*, 85:735–747, 2003.

[51] O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003a.

[52] O. Ledoit and M. Wolf. Honey i shrunk the sample covariance matrix. *Journal of Portfolio Management*, 31:603–621, 2003b.

[53] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*. 88:365–411, 2004.

[54] R. Levy. On the short-term stationarity of beta coefficients. *Financial Analysts Journal*, 27:55–62, 1971.

[55] J. Longerstaey and M. Spencer. *RiskMetrics - technical document.* Morgan Guaranty Trust Company of New York, New York, 4th edition, 1996.

[56] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. The accuracy of extrapolation (time-series) methods: results of a forecasting competition. *Journal of Forecasting*, 1:111–153, 1982.

[57] S. Makridakis, C. Chatfield, M. Hibon, T. Mills, K. Ord, and L. Simmons. The m2-compeition: a real time judgementally based forecasting study. *International Journal of Forecasting*, 9:5–22, 1993.

[58] S. Makridakis and M. Hibon. The m3-compeition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476, 2000.

[59] H. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[60] R.O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45:31-42, 1989.

[61] T.J. Moskowitz. An analysis of covariance risk and pricing anomalies. *Review of Financial Studies*, 16:417–457, 2003.

[62] R.J. Muirhead. Developments in eigenvalue estimation. In A.K. Gupta, editor, *Advances in Multivariate Statistical Analysis*, pages 277–288. D. Reidel Publishing Company, Holland, 1987.

[63] V.R. Ng, R.F. Engle, and M. Rothschild. A multi dynamic factor model for stock returns. *Journal of Econometrics*, 52:245–266, 1992.

[64] T.A. Oigard, A. Hanssen, and R.E. Hansen. The multivariate normal inverse gaussian distributionl em-estimation and analysis of synthetic aperture sonar data. Working paper, University of Tromso, 2004.

[65] S. Pafka and I. Kondor. Noisy covariance matrices and portfolio optimization ii. *Physica A: Statistical mechanics and its applications*, 319:487–494, 2003.

[66] S. Pafka and I. Kondor. Estimated correlation matrices and portfolio optimization. *Physica A: Statistical mechanics and its applications*, 343:623–634, 2004.

[67] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. A random matrix approach to cross-correlation in financial data. *eprint cond-mat0108023*, 2001.

[68] R. Roll and S. Ross. An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35:1073–1103, 1980.

[69] B. Rosenberg. Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis*, 9:263–274, 1974.

[70] B. Rosenow, V. Plerou, P. Gopikerishnan, L.A.N.Amaral, and H.E. Stanley. Application of random matrix theory to study cross-correlations of stock prices. *International Journal of Theoretical and Applied Finance*, 2:399–403, 2000.

[71] T.H. Rydberg. The normal inverse gaussian levy process: simulation and approximation. *Communications in Statistics - Stochastic Models*, 13:887–910, 1997.

[72] J. Schafer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4:Article 32, 2005.

[73] W. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9:277–293, 1963.

[74] C. Stein. Estimation of a covariance matrix. Rietz Lecture, 39th Annual Meeting IMS, Atlanta, GA, 1975.

[75] C. Stein. Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences*, 34:1373–1403, 1977.

[76] H. Theil. *Principles of Econometrics*. New York: John Wiley & Sons, Inc., 1971.

[77] O.A. Vasicek. A note on using cross-section information in bayesian estimation of security betas. *Journal of Finance*, 28:1233–1239, 1973.

[78] K. Winston. The 'efficient index' and prediction of portfolio variance. *Journal of Portfolio Management*, 19:27–34, 1993.

[79] R. Yang and J.O. Berger. Estimation of a covariance matrix using the reference prior. *Annals of Statistics*, 22:1195–1211, 1994.

[80] M.S. Yazici. *Forecasting the covariance structure of UK stock returns: test of a new model.* University of Warwick, Warwick Business School, 1996.